Microsoft Research

Microsoft

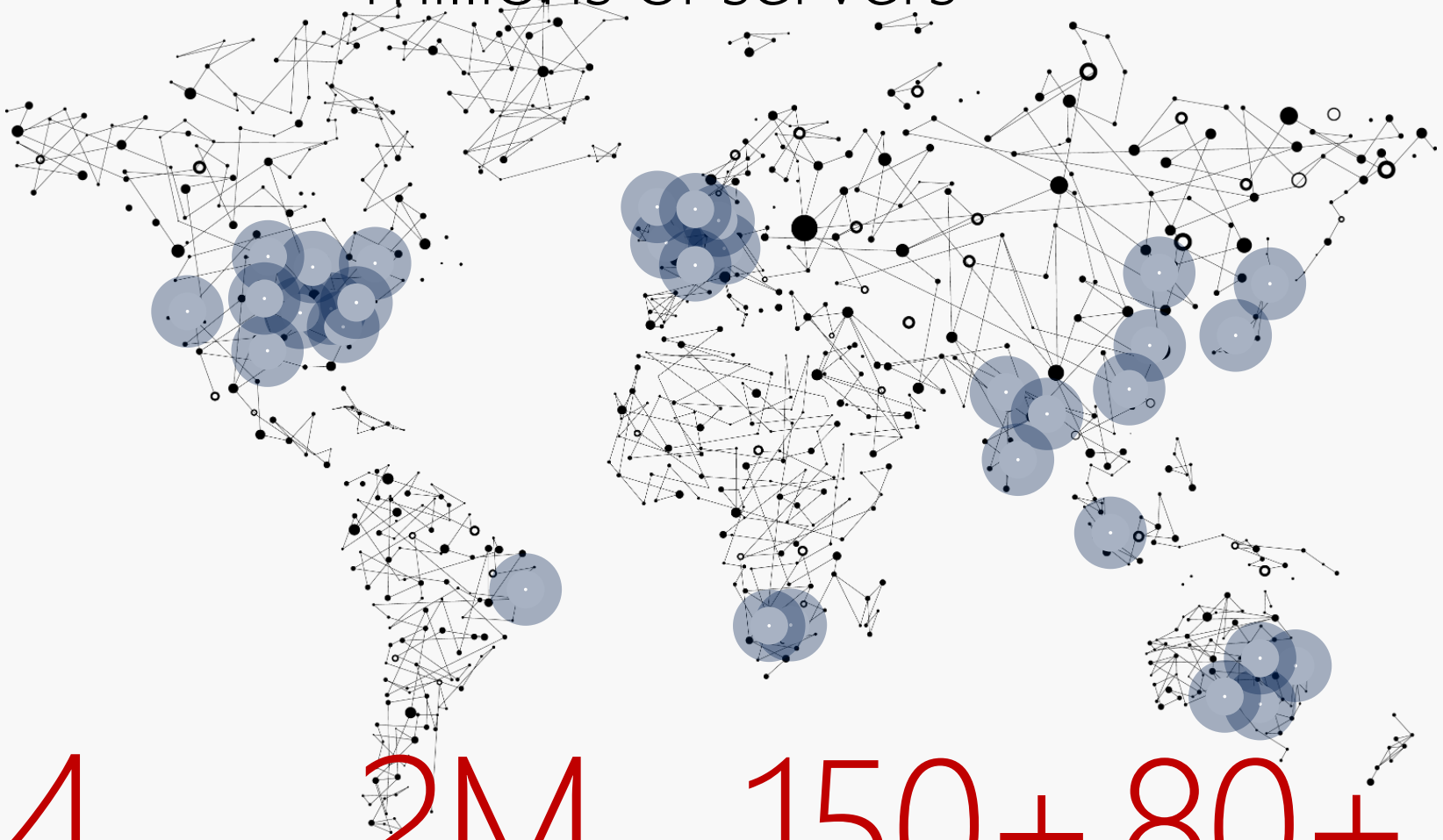
edge computing
a historical perspective & direction
10 years & counting

Victor Bahl
Distinguished Scientist
Director, Mobility & Networking Research
Microsoft Resea

Monday, August 20, 2018

Microsoft's big bet: Azure

millions of servers



54
Azure regions

2M
miles intra-DC fiber

150+
data centers

80+
Tb data on backbone

Microsoft's big bet: Azure

FY18: Intelligent Cloud: $23 billion revenue (Azure grew 89% in Q4)

FY17 growth numbers:

15%	97%	250+	2x
YoY Microsoft server products and cloud services revenue growth	YoY Azure Revenue Growth	FY17 Azure announcements	YoY Azure compute usage

>90%

of Fortune 500 use Microsoft Cloud

Microsoft's data centers

Columbia river, hydro-electric power

each facility is 8 MW in size, total of 64 MW

expanding rapidly, powered by wind farms

looking beyond cloud computing

October 29, 2008 in Bldg. 99

first paper

IEEE pervasive COMPUTING
MOBILE AND UBIQUITOUS SYSTEMS

VIRTUAL MACHINES

The Case for VM-Based Cloudlets in Mobile Computing

A new vision of mobile computing liberates mobile devices from severe resource constraints by enabling resource-inten applications to leverage cloud computing free of WAN dela jitter, congestion, and failures.

Total citations Cited by 2321

(as of 8/15/18)

first article

SHEPHERD'S Pi

Why a Cloudlet Beats the Cloud for Mobile Apps

Posted on December 13, 2009 by lewisshepherd

Microsoft

offloading & programming the edge (2009-10)

July 12 –14, 2009

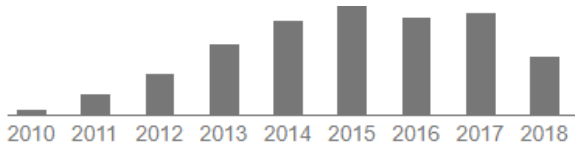
edge computing

MobiSys 2010

MAUI: Making Smartphones Last Longer with Code Offload

Eduardo Cuervo[†], Aruna Balasubramanian[‡], Dae-ki Cho[*],
Alec Wolman[§], Stefan Saroiu[§], Ranveer Chandra[§], Paramvir Bahl[§]
[†]Duke University, [‡]University of Massachusetts Amherst, [*]UCLA, [§]Microsoft Research

citation 1996 (as of 8/20/18)



ABSTRACT

This paper presents MAUI, a system that enables fine-grained energy-aware offload of mobile code to the infrastructure. Previous approaches to these problems either relied heavily on programmer support to partition an application, or they were coarse-grained requiring full process (or full VM) migration. MAUI uses the benefits of a managed code environment to offer the best of both worlds:

Given the t
the energy
foremost ch
One popu
vices is *remote execution*: applications can take advantage of the resource-rich infrastructure by delegating code execution to remote servers. For the last two decades, there have been many attempts to make mobile devices use remote execution to improve perfor-

Microsoft

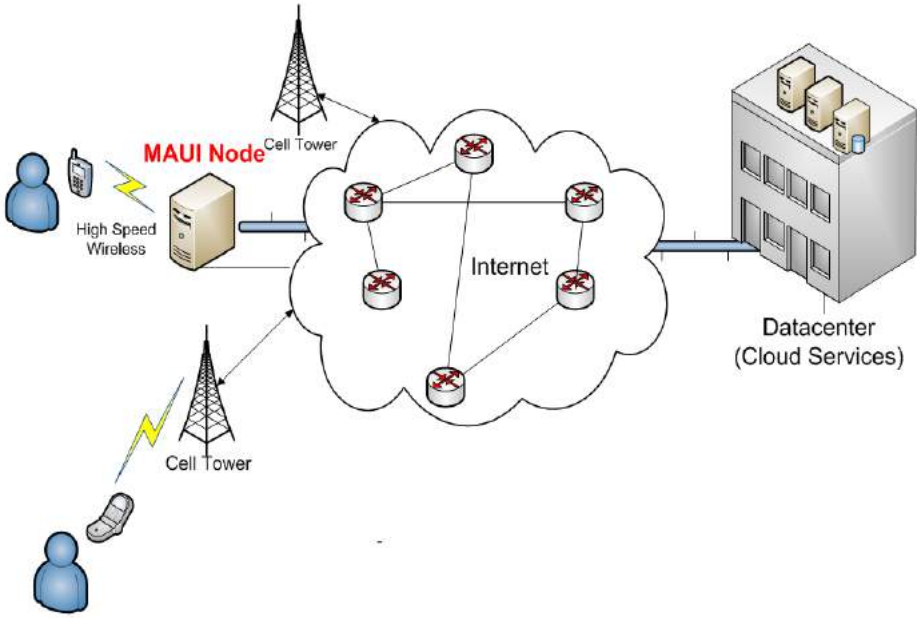
opportunistic use of infrastructure for dynamic offloading

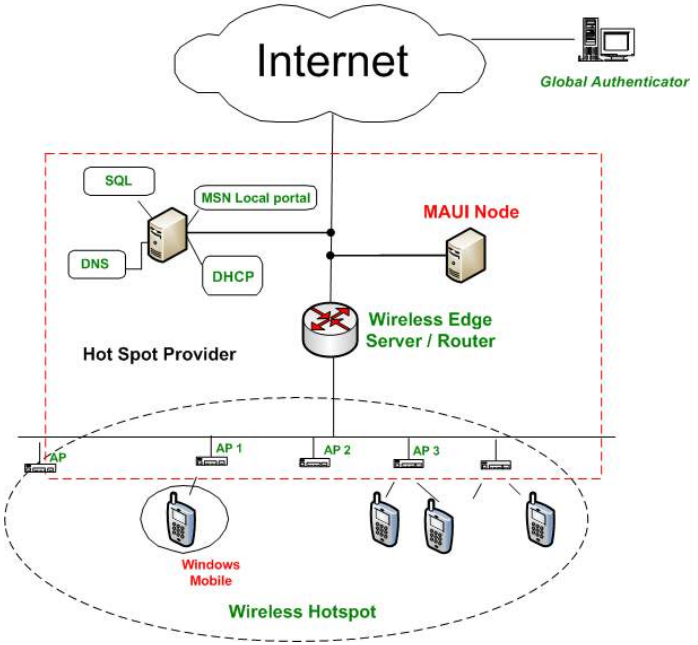
approach

- developers build standalone apps with simple annotations but *no changes to program logic*
- system uses nearby and cloud-server resources in *opportunistic manner*

properties

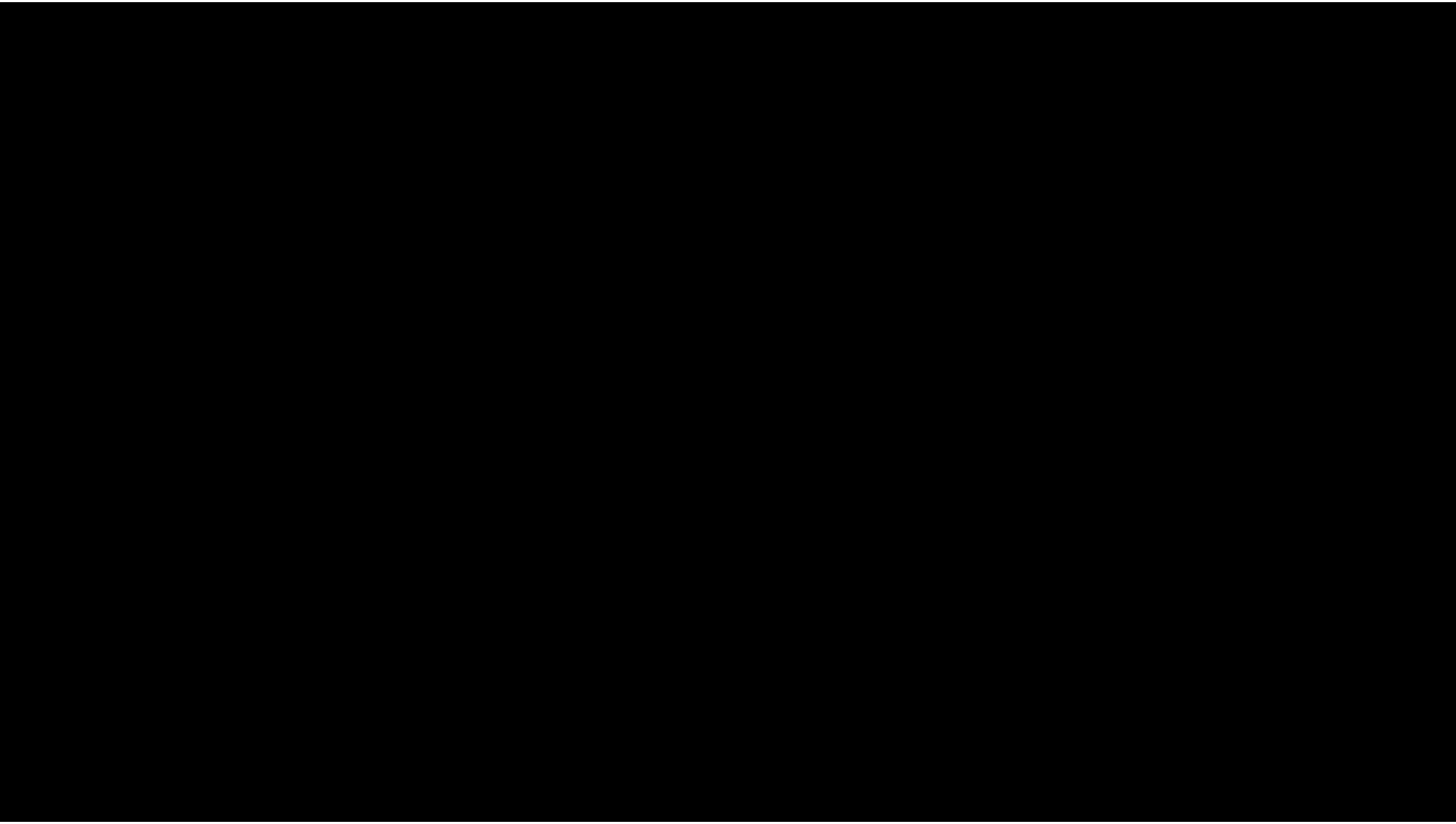
- apps. always work, even when disconnected
- simple programming model (lowers barrier to widespread adoption)





impact of latency on recognition performance

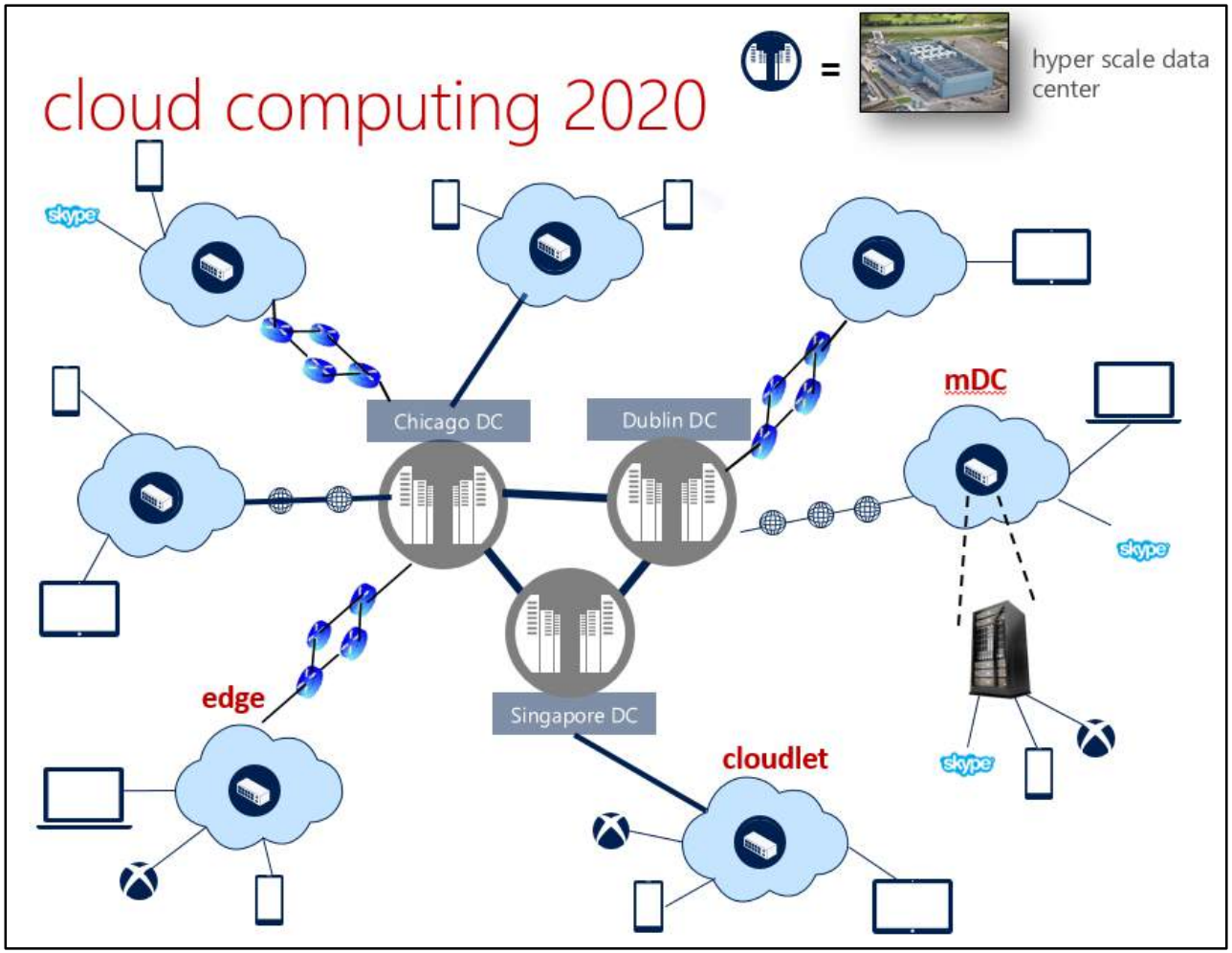
impact of latency on recognition performance



led to research, papers, keynotes, & a prediction

Dec. 12, 2013

slide 54



the disaggregated cloud!

Microsoft

prediction was based on

Dec. 12, 2013

Dec. 12, 2013

slide 52

the virtues of edge computing

latency reduction
- serve content immediately
- SSL termination with split TCP

bandwidth saving
- compression
- procrastination
- edge analytics

service & internet monitoring

reliable connectivity
- overlay networking
- path diversity

battery saving
- computation offloads
- client proxying

high-end game streaming
- lower device cost
- reduce developer fragmentation

new services & applications

protection against DoS

reduced load on DCs

there is plenty of research literature (incl. MSR's) that shows edge computing significantly enhances mobile experience

Microsoft

several developments since then

press articles

NETWORKWORLD **Microsoft researcher: Why Micro Datacenters really matter to mobile's future**

research projects

Elijah Cloudlet-based Mobile Computing 

Gabriel Wearable Cognitive Assistance using cloudlets

Government initiatives

NSF Workshop on Grand Challenges in Edge Computing

Workshop Dates | October 26, 2016

Venue | Washington, DC

standards

ETSI

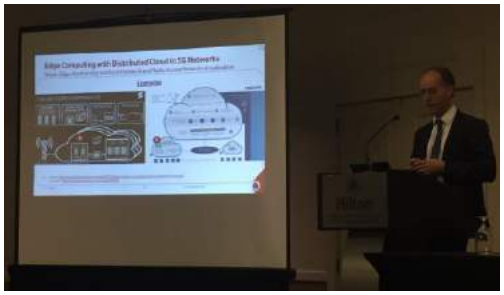
Mobile Edge Computing

Introduction | Our Role & Activities | Specifications | Blog

conferences

SEC 2016 **The First IEEE/ACM Symposium on Edge Computing**

October 27-28, 2016, Washington DC, USA



industry initiatives

Open Fog Consortium

OUR MISSION: TO DRIVE INDUSTRY AND ACADEMIC LEADERSHIP IN FOG COMPUTING ARCHITECTURE, TESTBED DEVELOPMENT, AND A VARIETY OF INTEROPERABILITY AND COMPOSABILITY DELIVERABLES THAT SEAMLESSLY LEVERAGE CLOUD AND EDGE ARCHITECTURES TO ENABLE END-TO-END IOT SCENARIOS.

Microsoft

… but we needed a killer app

MCS 2013

Vision: Cloud-Powered Sight for All

Showing the Cloud What You See

Paramvir Bahl Matthai Philipose Lin Zhong

Microsoft Research
Redmond, WA 98052

{bahl,matthaip,t-lizhon}@microsoft.com

ABSTRACT

We argue that for computers to do more for us, we need to show the cloud what we see and embrace cloud-powered sight for mobile users. We present sample applications that will be empowered by this vision, discuss why the timing is right to tackle it, and offer our initial thoughts on some of the important research challenges.

Categories and Subject Descriptors

A.1 [**General Literature**]: introductory and survey

General Terms

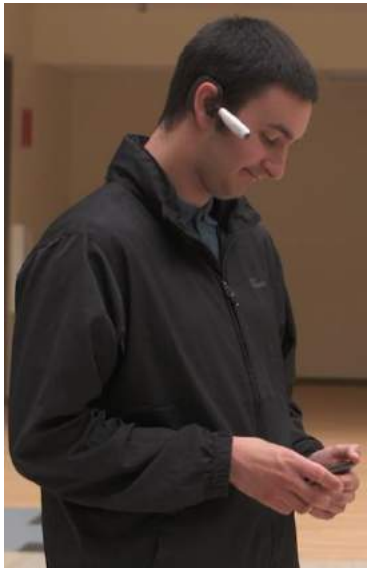
Algorithms, Design, Human Factors, Languages, Performance, Security

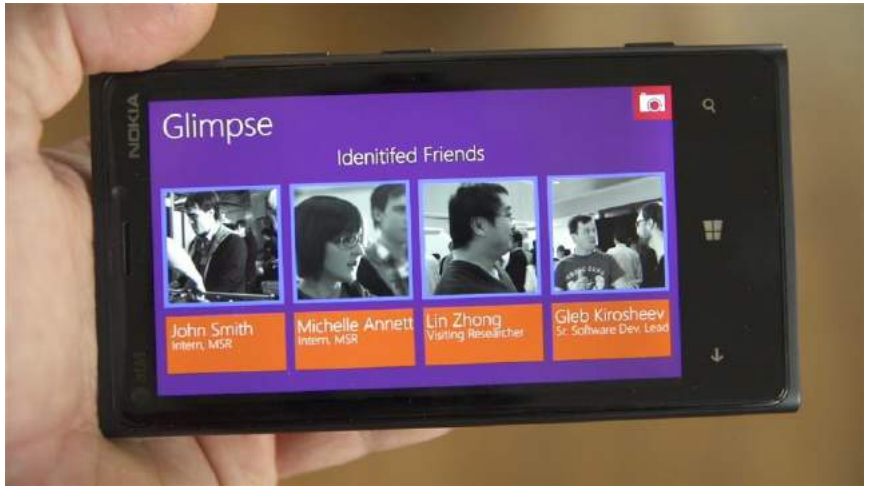
and share it with the cloud. Inspiring applications and services have been demonstrated that analyze continuously collected accelerometer and microphone data, e.g., [LPL+09], and occasionally data from phone camera, mation transfer is greater t limited.

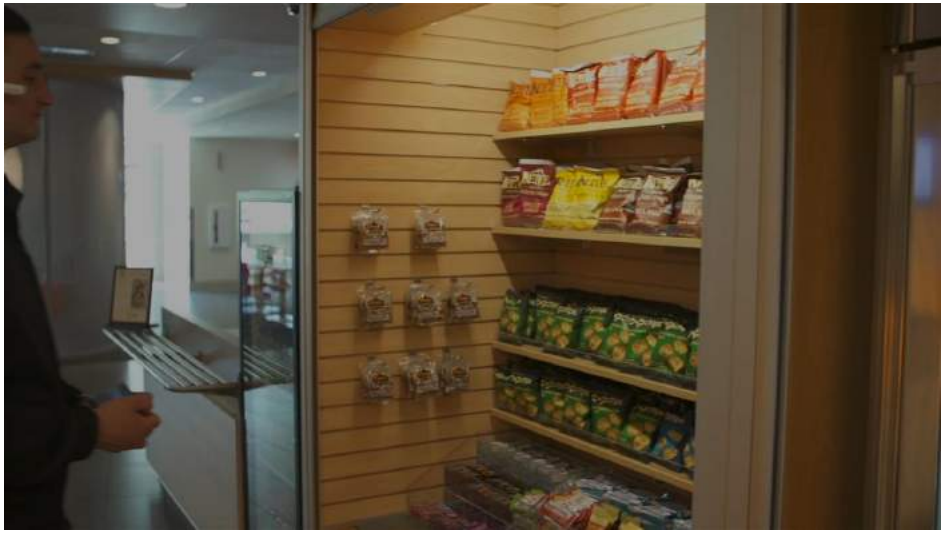
In this paper, we assert that should be configured to analyz to them at a much higher rate unleash the creativity of deve plications that make use of th cient. Today, we let the comp where and how we move, but *our computer see what we see*

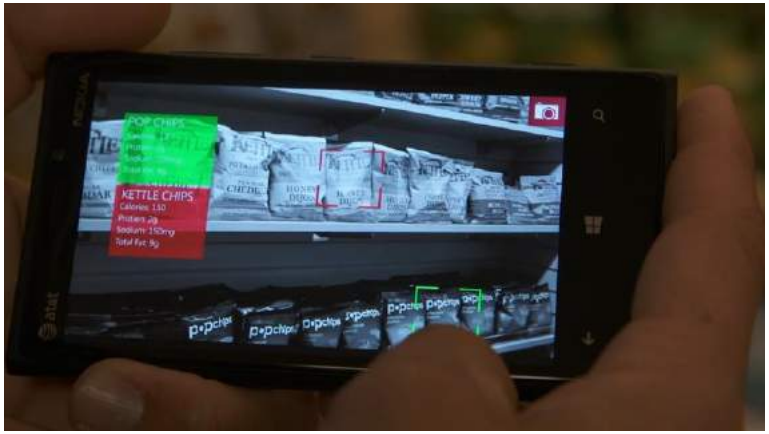
Microsoft

MSR's Glimpse project









Microsoft

MCS 2013

Vision: Cloud-Powered Sight for All

Showing the Cloud What You See

Paramvir Bahl Matthai Philipose Lin Zhong

MobiSys 2013

Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision

Robert LiKamWa[†,‡], Bodhi Priyantha[‡], Matthai Philipose[‡], Lin Zhong[†,‡], and Paramvir Bahl[‡]

best paper award

SenSys 2014

Glimpse: Continuous, Real-Time Object Recognition [illegible] Mobile Devices

Tiffany Yu-Han Chen
MIT CSAIL
yuhan@csail.mit.edu

Lenin Ravindranath
Microsoft Research
lenin@microsoft.com

Shuo Deng
MIT CSAIL
shuodeng@csail.mit.edu

Paramvir Bahl
Microsoft Research
bahl@microsoft.com

Hari Balakrishnan
MIT CSAIL
hari@csail.mit.edu

ABSTRACT

Glimpse is a continuous, real-time object recognition system for camera-equipped mobile devices. Glimpse captures full-motion video, locates objects of interest, recognizes and labels them, and tracks them from frame to frame for the user. Because the algorithms for object recognition entail significant computation, Glimpse runs them on server machines. When the latency between the server and mobile device is higher than a frame-time, this approach lowers object-recognition accuracy. To regain accuracy, Glimpse uses an *active cache* of video frames on the mobile device. A subset of the frames in the active cache are used to track objects on the mobile, using (stale) hints about objects that arrive from

Frame 1 (t = 0 ms) Frame 20 (t = 660 ms)

Figure 1: Offloading every frame to a server reduces trackability (right): the stop sign's location is wrong.

sigmobile

highlights

Microsoft

canonical example for edge computing

the connected car

Microsoft

in–vehicle video analytics for detecting open parking spaces in urban environment

Giulio

Microsoft Research

UPMC SORBONNE UNIVERSITÉS

UCL

January 2015

Microsoft

MCS 2013

Vision: Cloud-Powered Sight for All

Showing the Cloud What You See

Paramvir Bahl Matthai Philipose Lin Zhong

MobiSys 2013

Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision

Robert LiKamWa[†,‡], Bodhi Priyantha[‡], Matthai Philipose[‡], Lin Zhong[†,‡], and Paramvir Bahl[‡]

best paper award

SenSys 2014

Glimpse: Continuous, Real-Time Object Recognition [illegible] Mobile Devices

Tiffany Yu-Han Chen Lenin Ravindranath Shuo Deng

sigmobile

highlights

SEC 2017

ParkMaster: An in–vehicle, edge–based video anal[illegible] service for detecting open parking spaces in urban environments

Giulio Grassi
Sorbonne Universités, UPMC, LIP6

Paramvir Bahl
Microsoft Research, Redmond

Kyle Jamieson
Princeton University; University College London

Giovanni Pau
Sorbonne Universités, UPMC, LIP6

ABSTRACT

We present the design and implementation of ParkMaster, a system that leverages the ubiquitous smartphone to help drivers find parking spaces in the urban environment. ParkMaster estimates parking space availability using video gleaned from drivers' dash-mounted smartphones on the network's edge, uploading analytics about the street to the cloud in real time as participants drive. Novel lightweight parked-car localization algorithms enable the system to estimate each parked car's approximate location by fusing information from phone's camera, GPS, and inertial sensors, tracking and count-

KEYWORDS

Fog computing, Edge computing, Mobile Systems, Visual Analytics

1 INTRODUCTION

Urban driving can be challenging and stressful, with the task of searching for parking spaces one of the key reasons. For example, a 2007 study in San Francisco [45] shows that in one of that city's commercial districts a driver spends on average 6.5 minutes to find a parking spot after reaching the

Microsoft

aha moment!

THE WALL STREET JOURNAL.
China's 100 Million Surveillance Cameras

theguardian
You're being watched: there's one CCTV camera for every 32 people in UK

a  for every 8 people in the US & for every 29 people worldwide!

→ live video streams are being generated from factory floors, traffic intersections, camera mounted on cars, police vehicles, & retail shops

extract value from video streams in-context, in-the-moment to generate actions & workflows

with cloud computing, it's the golden era for computer vision, AI & machine learning

potential to impact science, society & business

Microsoft

first attempt: public security

Aakanksha

security alerts, tracking

locating objects of interest

crowd Analytics & managment

prevailing approach (at the time):

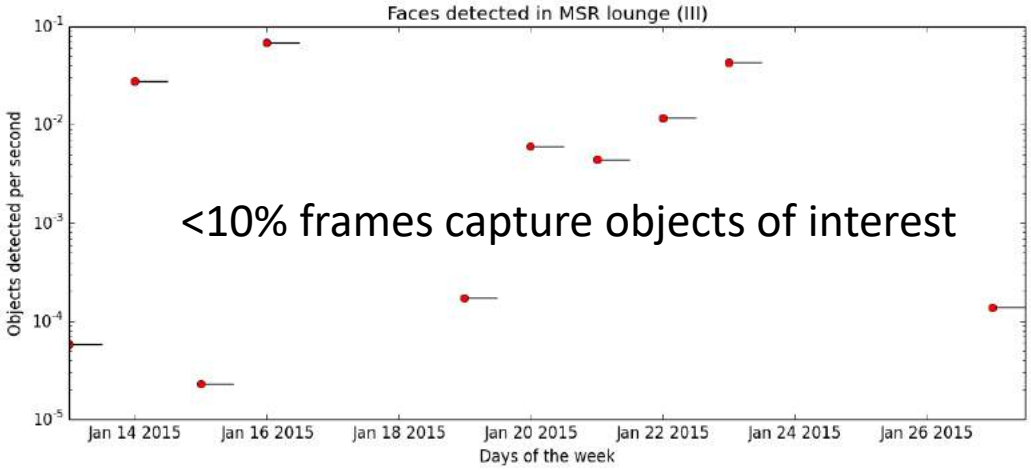
upload video to the cloud for remote (offline) analysis

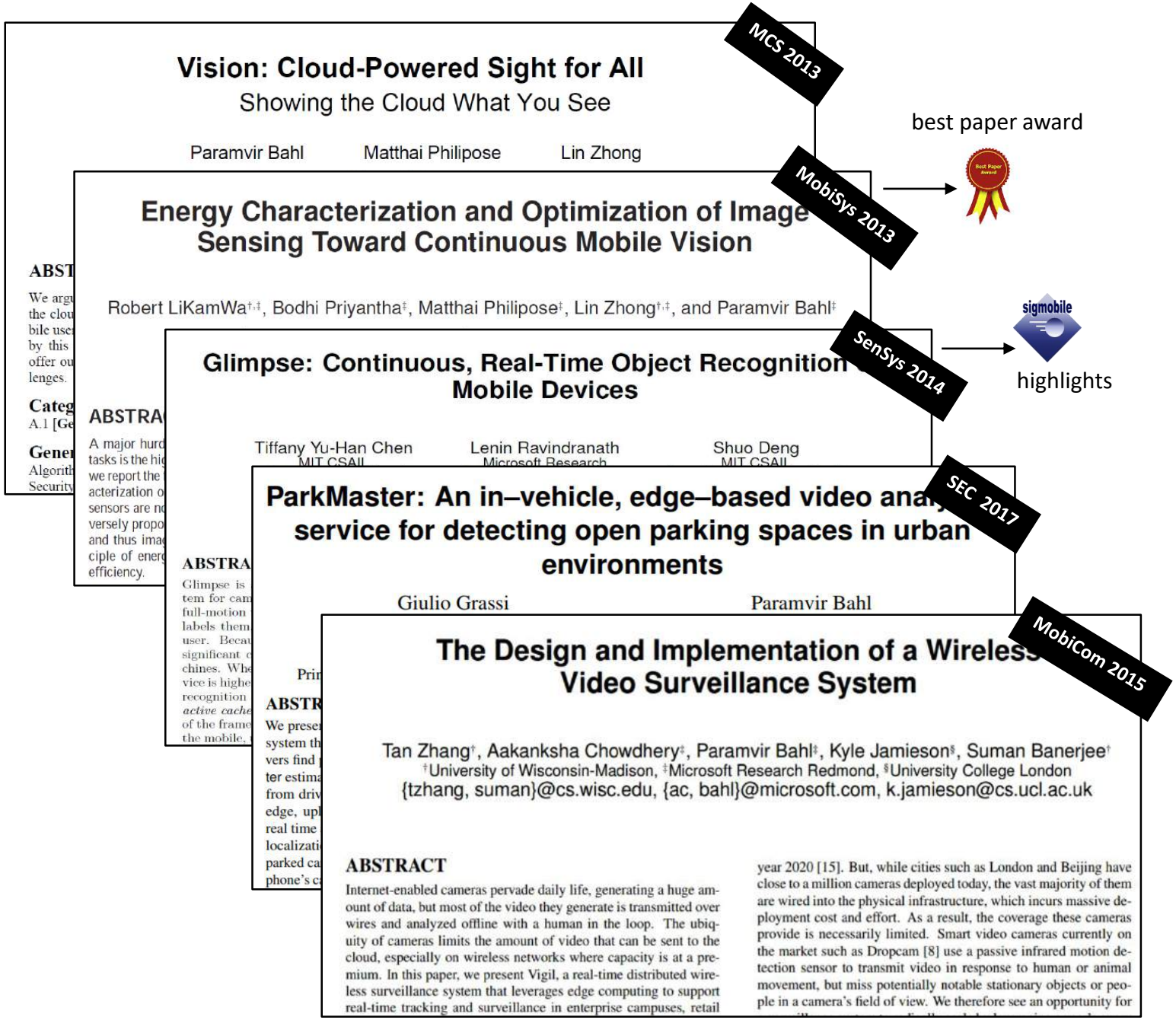
limitations

- large quantities of data (>10GB/hour)
- bandwidth availability limited coverage & accuracy
- human availability limited the systems usefulness
  - no automatic real-time tracking or alerts

Microsoft

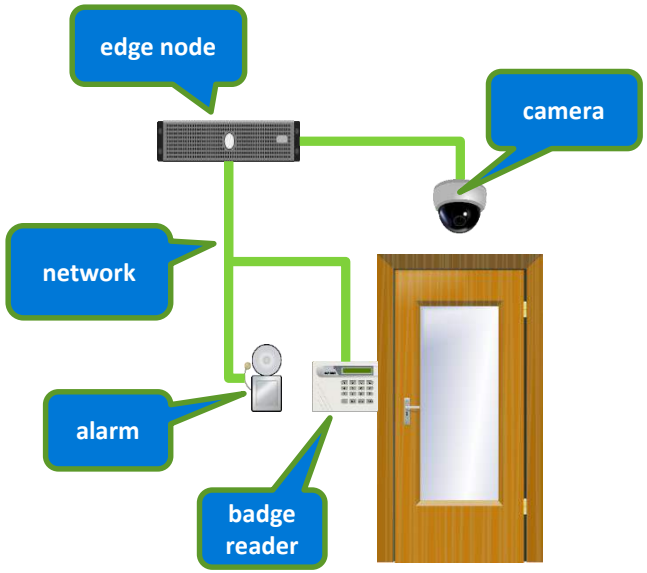
saving network bandwidth
(increasing coverage & accuracy)




MCS 2013
Vision: Cloud-Powered Sight for All
Showing the Cloud What You See
Paramvir Bahl
Matthai Philipose
Lin Zhong
best paper award
MobiSys 2013
Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision
Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl
sigmobile
SenSys 2014
highlights
Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices
Tiffany Yu-Han Chen
Lenin Ravindranath
Shuo Deng
SEC 2017
ParkMaster: An in–vehicle, edge–based video analytics service for detecting open parking spaces in urban environments
Giulio Grassi
Paramvir Bahl
MobiCom 2015
The Design and Implementation of a Wireless Video Surveillance System
Tan Zhang, Aakanksha Chowdhery, Paramvir Bahl, Kyle Jamieson, Suman Banerjee
University of Wisconsin-Madison, Microsoft Research Redmond, University College London
{tzhang, suman}@cs.wisc.edu, {ac, bahl}@microsoft.com, k.jamieson@cs.ucl.ac.uk
ABSTRACT
Internet-enabled cameras pervade daily life, generating a huge amount of data, but most of the video they generate is transmitted over wires and analyzed offline with a human in the loop. The ubiquity of cameras limits the amount of video that can be sent to the cloud, especially on wireless networks where capacity is at a premium. In this paper, we present Vigil, a real-time distributed wireless surveillance system that leverages edge computing to support real-time tracking and surveillance in enterprise campuses, retail
year 2020 [15]. But, while cities such as London and Beijing have close to a million cameras deployed today, the vast majority of them are wired into the physical infrastructure, which incurs massive deployment cost and effort. As a result, the coverage these cameras provide is necessarily limited. Smart video cameras currently on the market such as Dropcam [8] use a passive infrared motion detection sensor to transmit video in response to human or animal movement, but miss potentially notable stationary objects or people in a camera's field of view. We therefore see an opportunity for

Microsoft

fun project: securing corporate buildings



Microsoft

some disturbing local news

Hit-and-run driver nearly kills woman on bike in Bellevue

BY KOMO NEWS | WEDNESDAY, MARCH 23RD 2016

ADVERTISEMENT

Car strikes, kills toddler in stroller in Bellevue

Originally published September 29, 2015 at 11:03 am | *Updated September 30, 2015 at 10:27 am*

77-year-old pedestrian killed by teen driver in Bellevue

BY TIM HAECK, KIRO Radio Reporter | December 1, 2014 @ 10:17 am

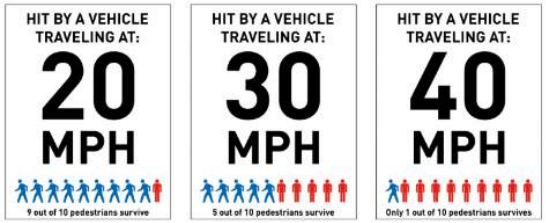
local TV coverage

impact of crashes (2010): economic cost: $242B; societal harm: $836B (*source: NHTSA*)

Microsoft

traffic safety: a world-wide movement

- 1.2 million people die on the world's roads every year
- 20-50 million suffer non-fatal injuries
- in the US, 19,000 people were killed in the first 6 months of 2016 (up 9% compared to 2015)



WIKIPEDIA
The Free Encyclopedia

Vision Zero is a multi-national road traffic safety project that aims to achieve a highway system with no fatalities or serious injuries in road traffic. It started in Sweden and was approved by their parliament in October 1997.

Microsoft

cities all over North America are embracing it

VISION ZERO
ANCHORAGE

Vision Zero
City of Boston

VISION ZERO
LOS ANGELES | 2015-2025

Vision Zero
Calgary

VISION ZERO
CHICAGO

Vision Zero San Jose

VZ SF 2024 VISION ZERO SF

VISION ZERO
Zero Traffic Deaths in San Diego by 2025

VISION ZERO
SAFE STREETS FOR WASHINGTON, DC

VISION ZERO
SANTA BARBARA

city planners care about -

- how often are vehicles speeding & failing to yield to  ?
- are pedestrians disregarding traffic signals?
- are bicyclists ignoring 

or are they running  ?
- any trends that hint at the reasons why certain 

are broken in certain places?
- did a countermeasure have the desired effect?

courtesy: Franz Loewenherz, Senior Transportation Planner, City of Bellevue, WA

city planners need data & analytics to perform corrective measures

2005 - 2010 60 collisions recorded by the Bellevue Police Department

→

In 2013, WSDOT built a new roundabout at the intersection

Microsoft

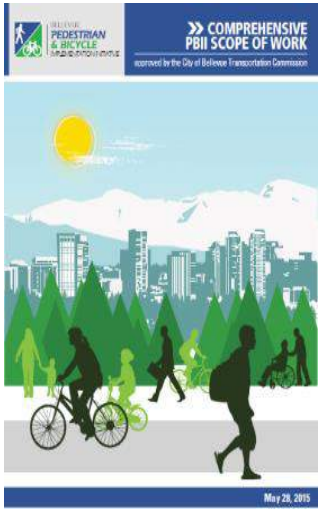
…we got going, we had a “killer” application and it was about saving lives

Bellevue, WA + Microsoft Research

Vision Zero: eliminate pedestrian/biker deaths

Use widely deployed traffic cameras

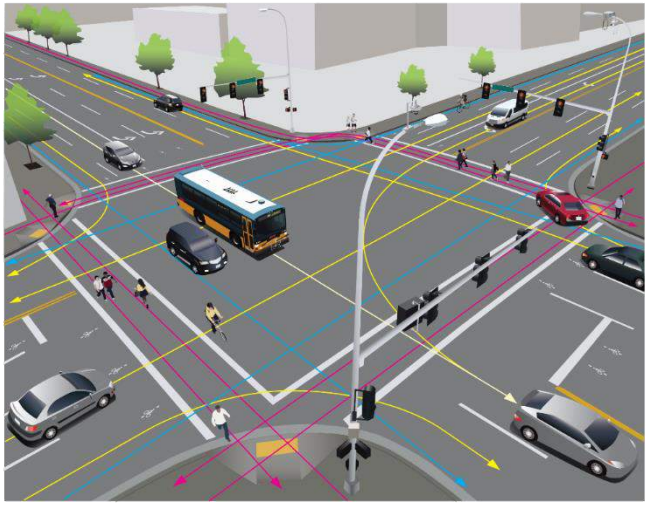
- Car/bike/ped counts, near-collisions, anomalies

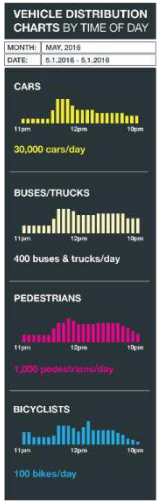


BELLEVUE
PEDESTRIAN
& BICYCLE
IMPLEMENTATION INITIATIVE

Making Bellevue a great place to walk and bike.

next-generation traffic control





Amy Carlson,
Vice President & Area Office Manager, CH2M Hill

Microsoft

picked up by local media

GeekWire

Microsoft looks to stop bike crashes before they happen, testing Minority Report-style predictive intelligence

BY LISA STIFFLER on October 14, 2015 at 1:00 pm

declined interview but…

"Microsoft, Bellevue team up to prevent crashes"

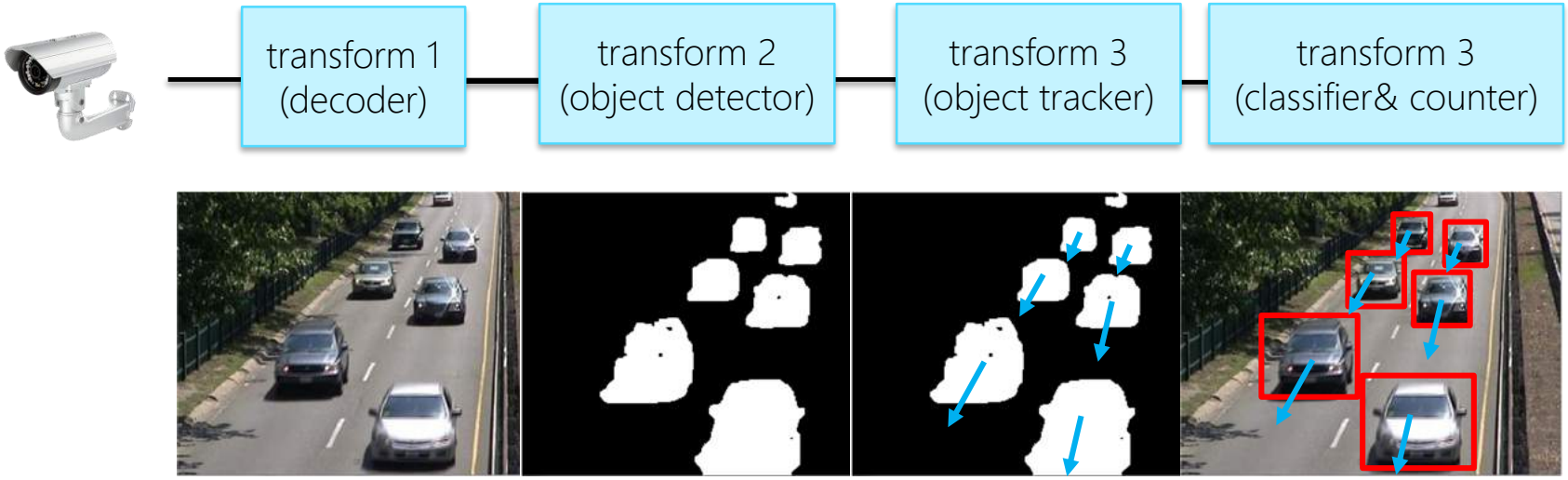
KIRO 7

Microsoft

video query: pipeline of transforms

vision algorithms ("*transforms*") chained together
transforms implement specified interfaces

example: count the number of moving cars on a road segment



Microsoft

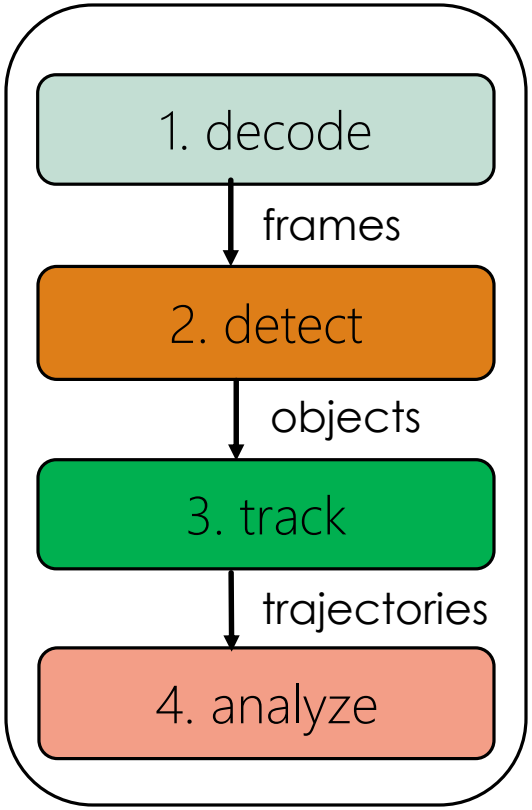
many implementation choices

40+ detector implementations

- motion-based: background subtraction
- DNN-based: Yolo detection
- exhaustive search

60+ tracker implementations

- moving pattern
- color histogram
- key-point features: SURF, SIFT



which implementation will you select?

Microsoft

which implementation is better?

DNN + histogram (0.17 fps

BGS + movement (42.3 fps)

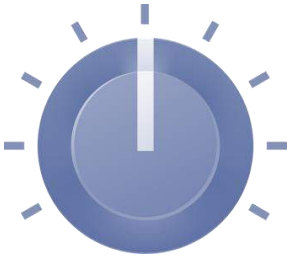
rosoft

each implementation's performance is impacted by the selection of "knob" positions

Haoyu

frame rate

30 fps for HD cameras



resolution

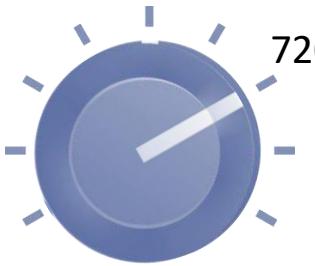
1080p, 720p, 480p…

window size

region of interest

Licence Plate Reader

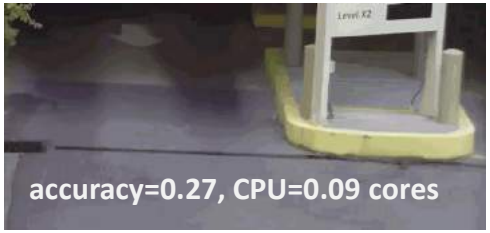
3



720p

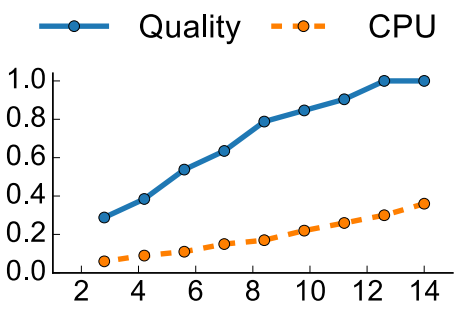
1

480p

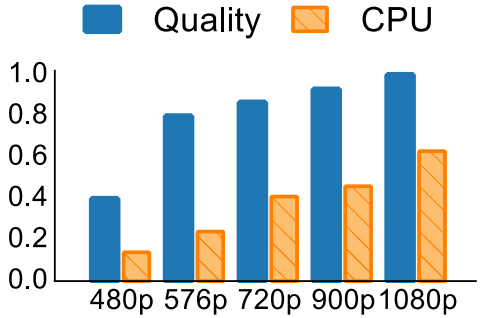


Microsoft

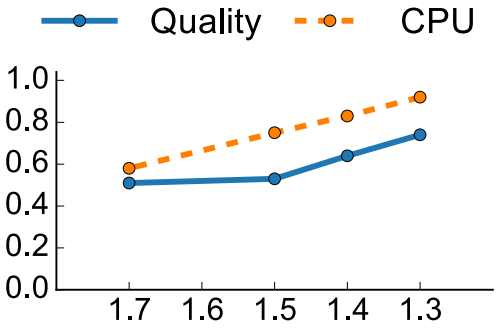
knobs/parameters impact quality & resource demands



frame rate



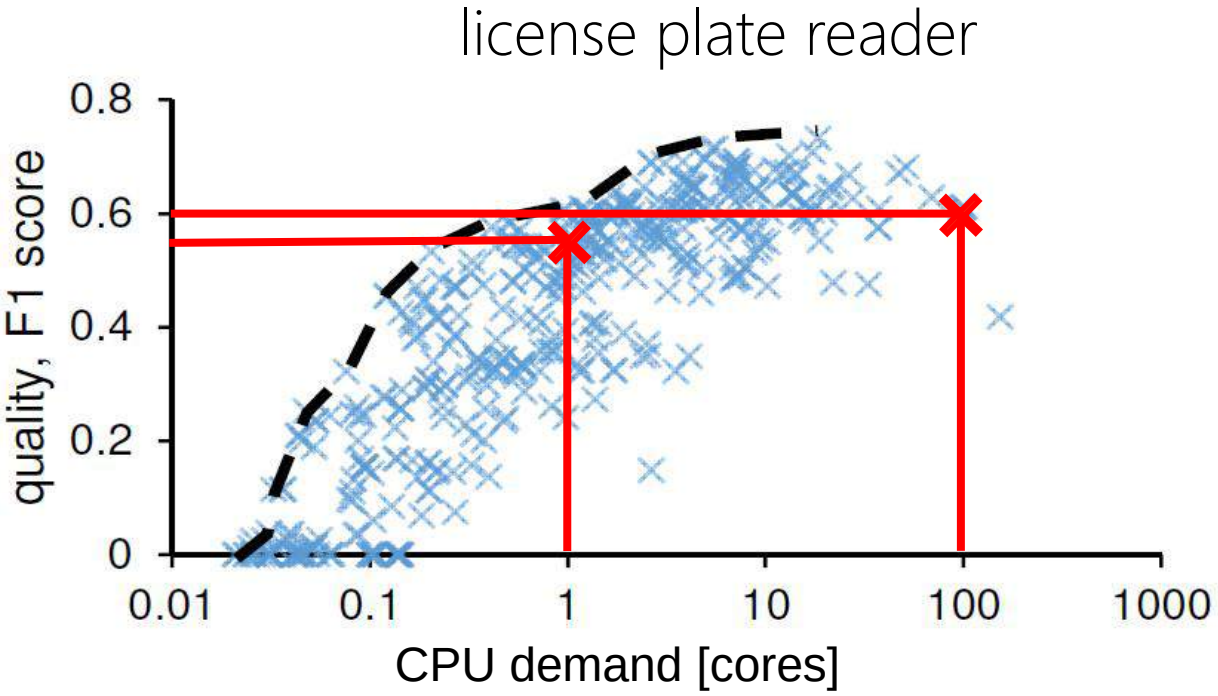
resolution



window size

Microsoft

impact of knobs/parameters on quality & resource demands

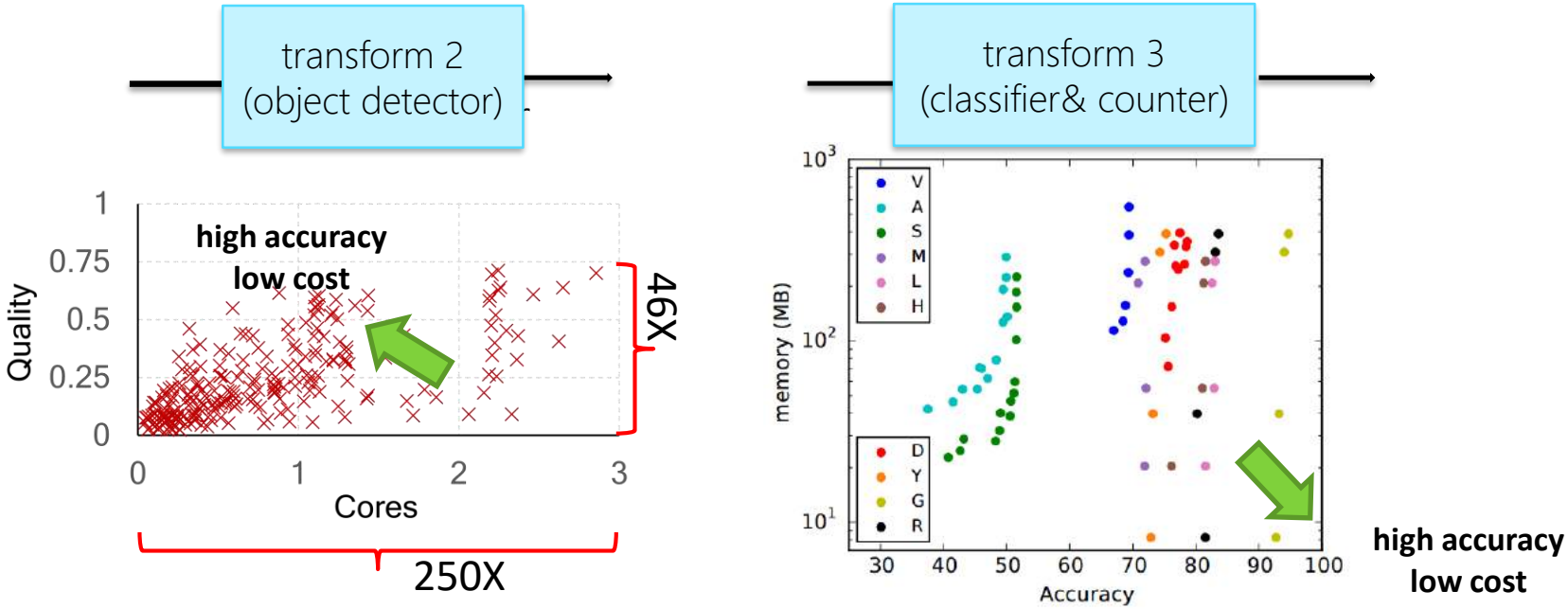


orders of magnitude cheaper resource demand for little quality drop

no analytical models to predict resource-quality tradeoff

Microsoft

resource - quality profile



best car tracker[1] — 1 fps on an 8-core CPU
DNN for object classification[2] — 30GFlops

no one plan is uniformly the best…
differ by 46x in their accuracy, 250x in speed!
best plan is dependent on the camera, lighting, track direction, object color, …

[1] VOT Challenge 2015 Results.
[2] Simonyan et al. CVPR abs/1409.1556, 2014

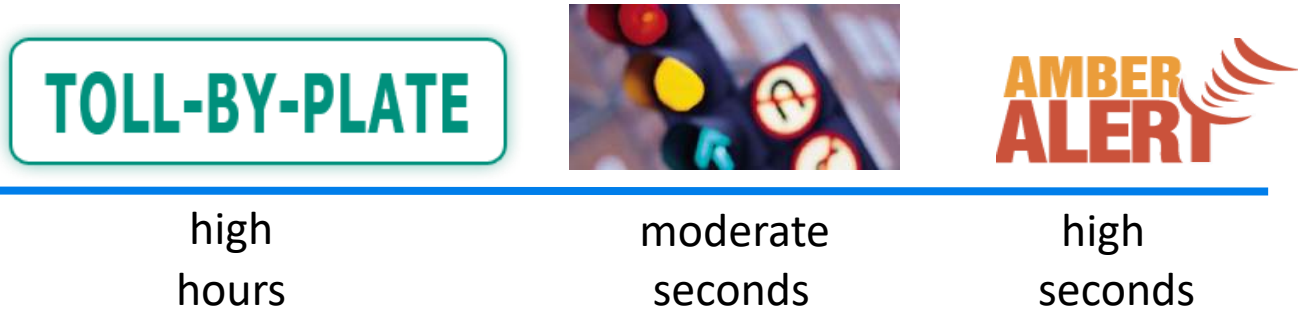
Microsoft

processing thousands of live streams

to support different types of queries at scale:

- must reduce processing cost of a query
- must schedule resources efficiently across queries

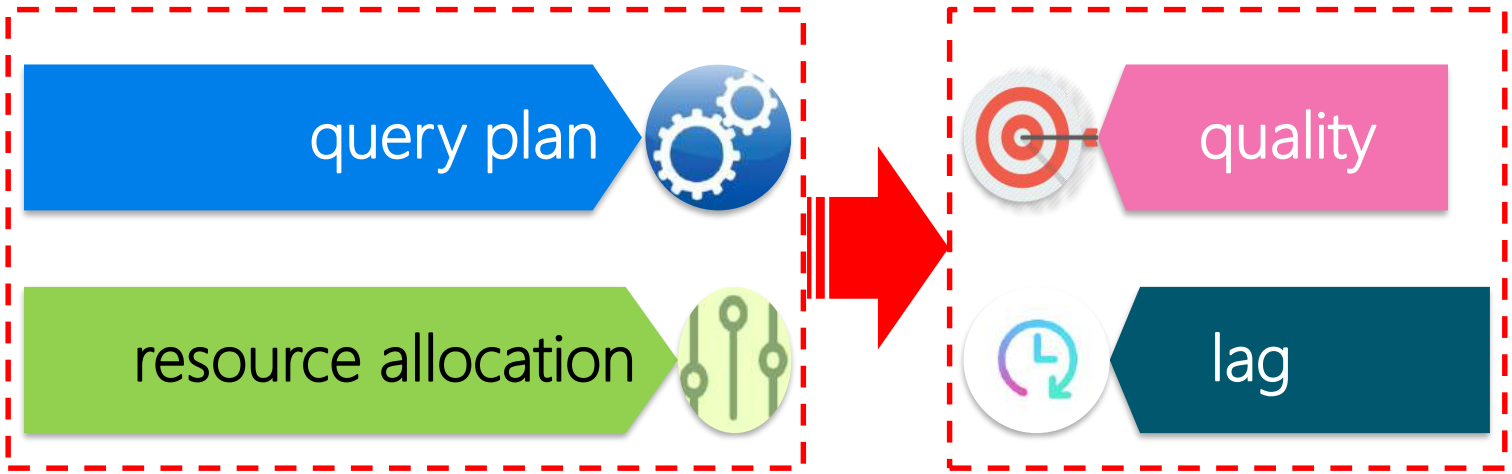
lag: time difference between frame arrival and frame processing



accuracy	high	moderate	high
lag	hours	seconds	seconds

Microsoft

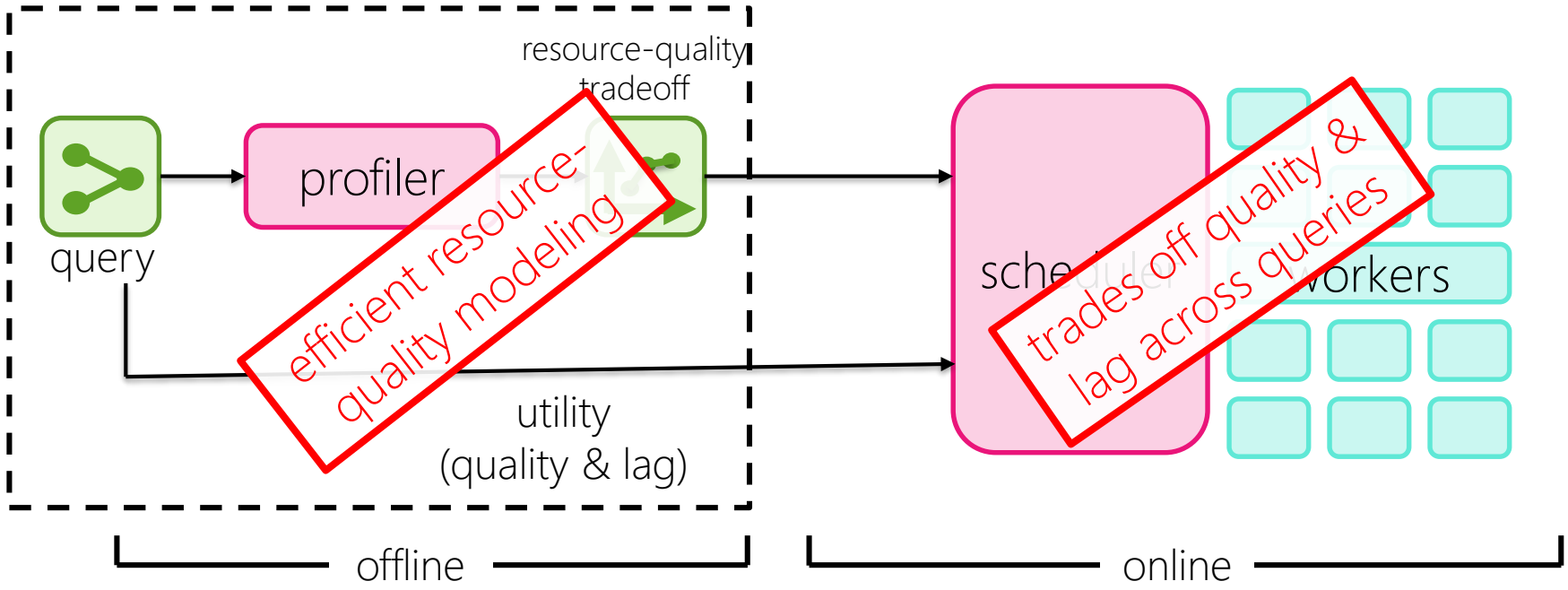
what is the best implementations for a video analytics query?

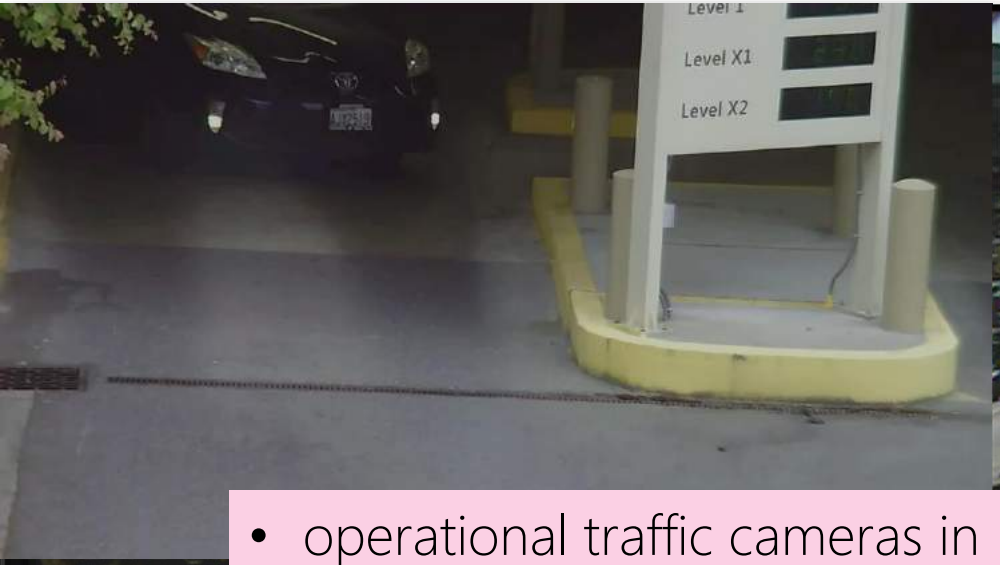


the configuration & resource allocation that maximizes quality & minimizes lag within the given resource capacity is the best implementation

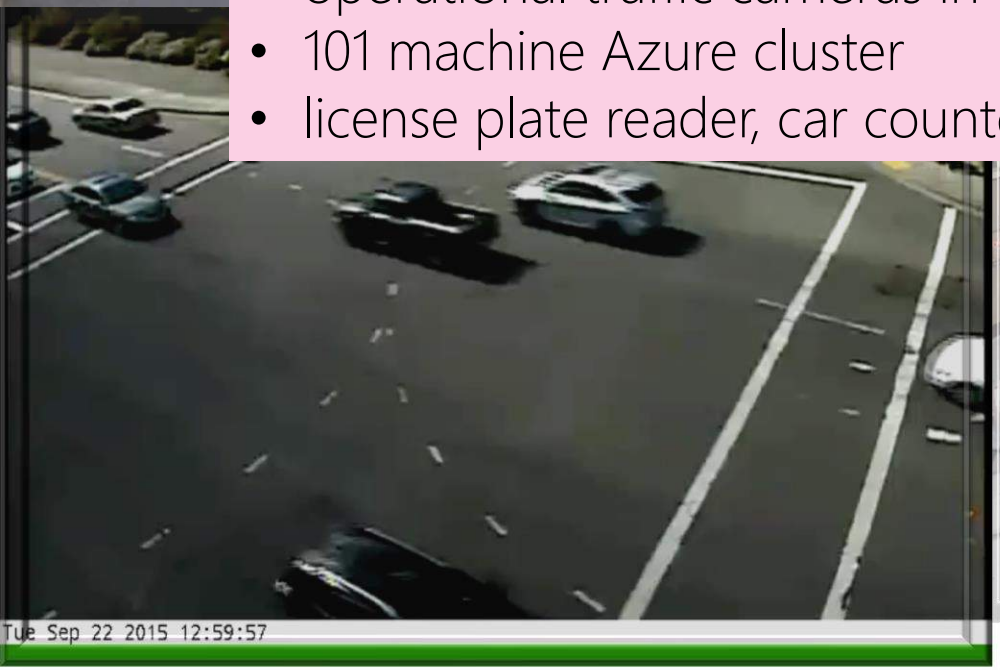
Microsoft

system design



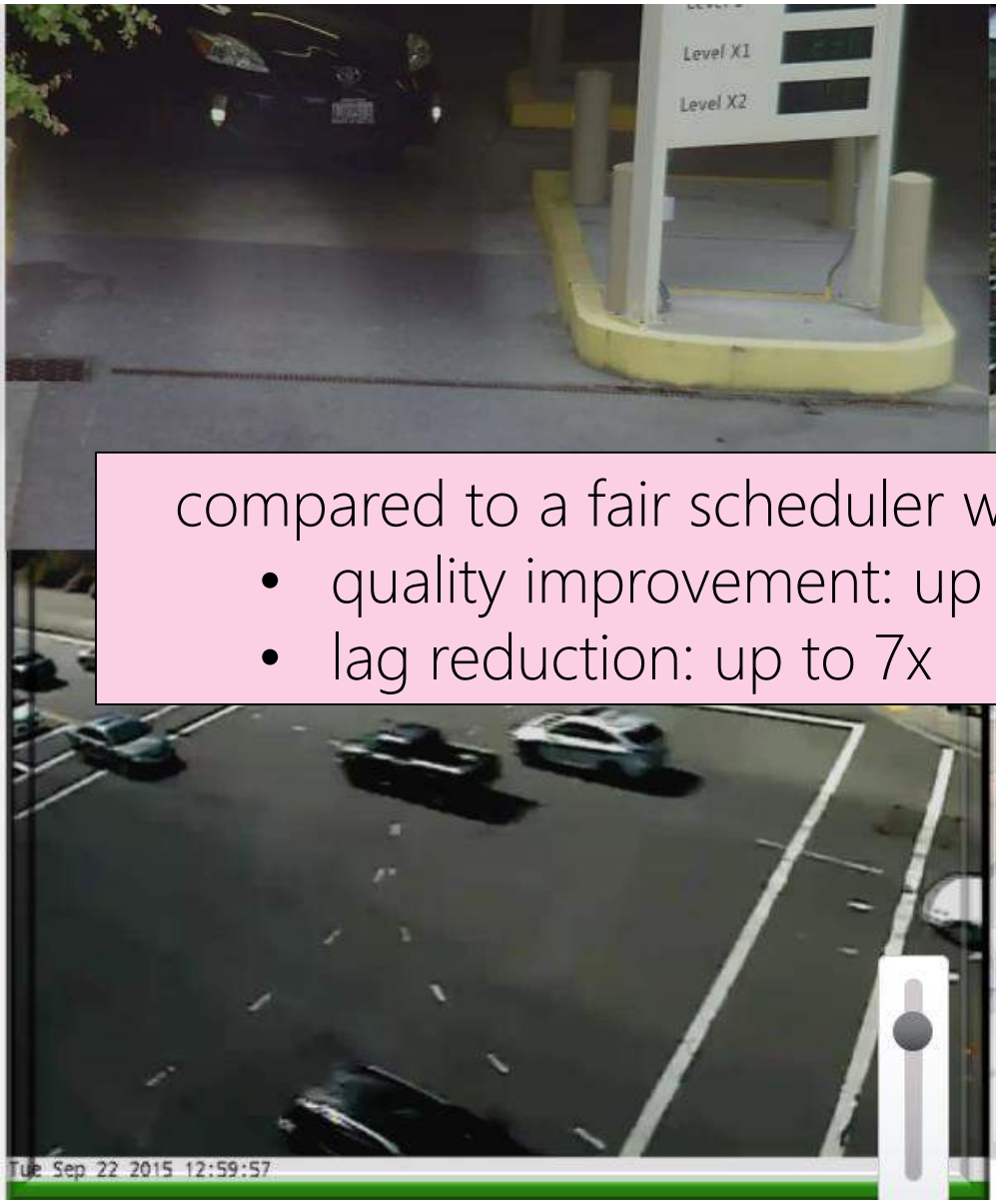


- operational traffic cameras in Bellevue and Seattle
- 101 machine Azure cluster
- license plate reader, car counter, DNN classifier, object tracker



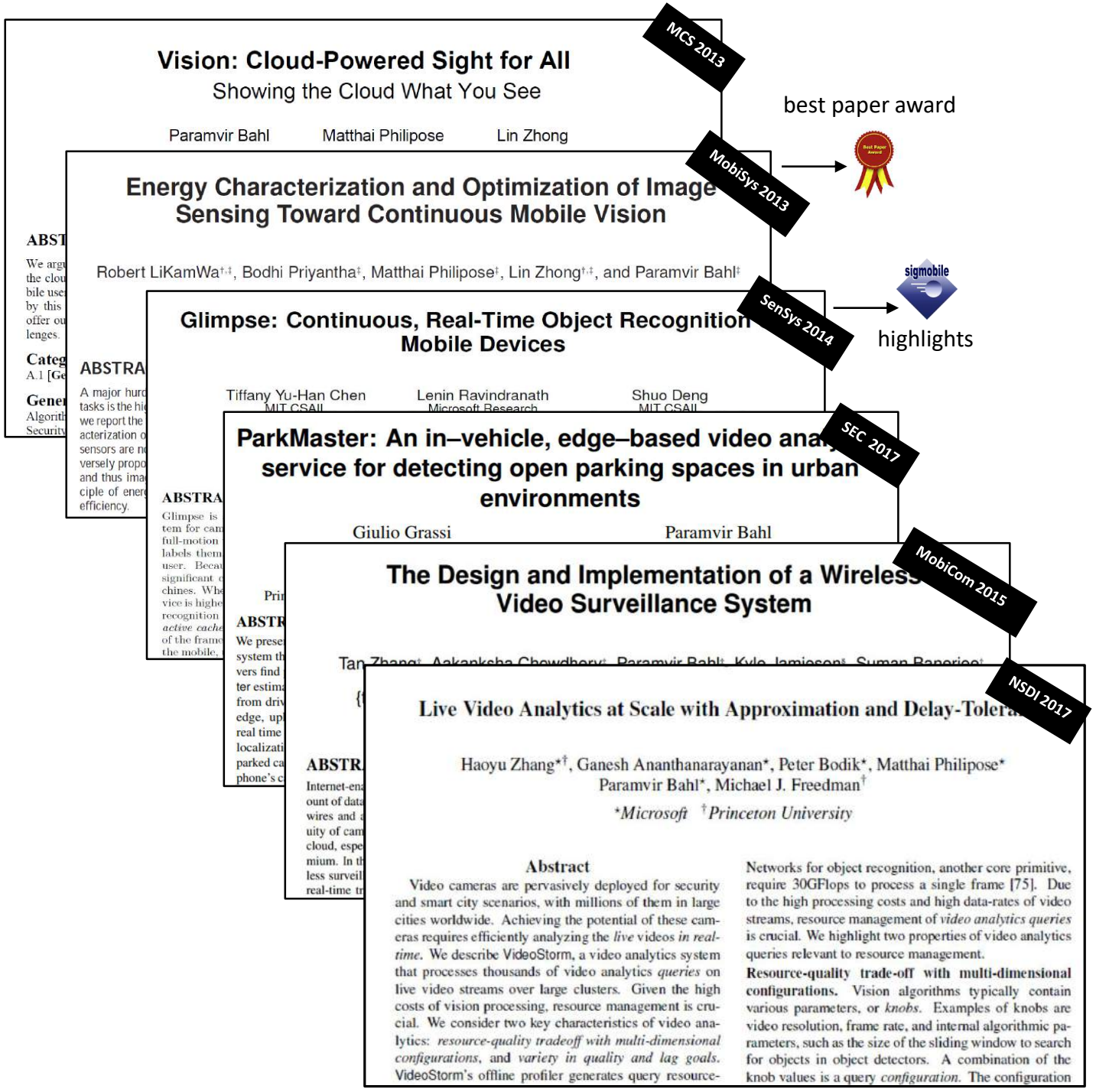
results

details in our NSDI 2017 paper



compared to a fair scheduler with varying burst duration:

- quality improvement: up to 80%
- lag reduction: up to 7x


Vision: Cloud-Powered Sight for All
Showing the Cloud What You See
Paramvir Bahl
Matthai Philipose
Lin Zhong
MCS 2013
best paper award
Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision
Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl
MobiSys 2013
Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices
Tiffany Yu-Han Chen
Lenin Ravindranath
Shuo Deng
SenSys 2014
sigmobile
highlights
ParkMaster: An in–vehicle, edge–based video analytics service for detecting open parking spaces in urban environments
Giulio Grassi
Paramvir Bahl
SEC 2017
The Design and Implementation of a Wireless Video Surveillance System
MobiCom 2015
Live Video Analytics at Scale with Approximation and Delay-Tolera
NSDI 2017
Haoyu Zhang*†, Ganesh Ananthanarayanan*, Peter Bodik*, Matthai Philipose*
Paramvir Bahl*, Michael J. Freedman†
*Microsoft †Princeton University
Abstract
Video cameras are pervasively deployed for security and smart city scenarios, with millions of them in large cities worldwide. Achieving the potential of these cameras requires efficiently analyzing the live videos in real-time. We describe VideoStorm, a video analytics system that processes thousands of video analytics queries on live video streams over large clusters. Given the high costs of vision processing, resource management is crucial. We consider two key characteristics of video analytics: resource-quality tradeoff with multi-dimensional configurations, and variety in quality and lag goals. VideoStorm's offline profiler generates query resource-
Networks for object recognition, another core primitive, require 30GFlops to process a single frame [75]. Due to the high processing costs and high data-rates of video streams, resource management of video analytics queries is crucial. We highlight two properties of video analytics queries relevant to resource management.
Resource-quality trade-off with multi-dimensional configurations. Vision algorithms typically contain various parameters, or knobs. Examples of knobs are video resolution, frame rate, and internal algorithmic parameters, such as the size of the sliding window to search for objects in object detectors. A combination of the knob values is a query configuration. The configuration

Microsoft

…and we have been deploying & learning

(Cambridge, U.K)

we recognized it as

when it really is

classified → / truth ↓	vehicles	bikes	peds	none
vehicle	**0.95**	0.01	0.02	0.02
bike	0.08	**0.67**	0.16	0.08
pedestrian	0.15	0.15	**0.73**	0.05
None	0.09	0.03	0.11	**0.81**

Microsoft

multi-tenancy

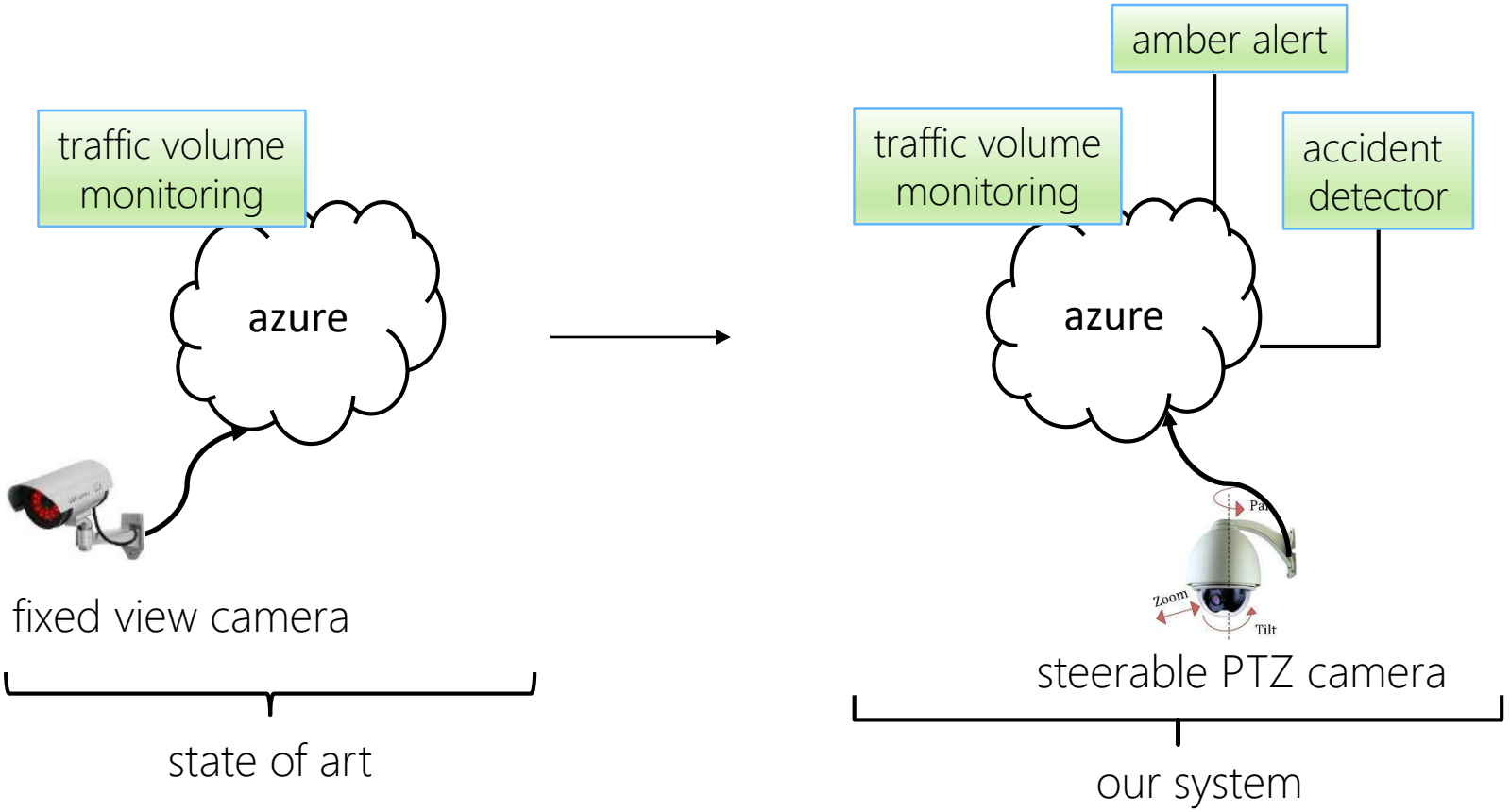
can a existing network of cameras be used by more than a single customer?

steerable cameras

servicing multiple applications simultaneously

Microsoft

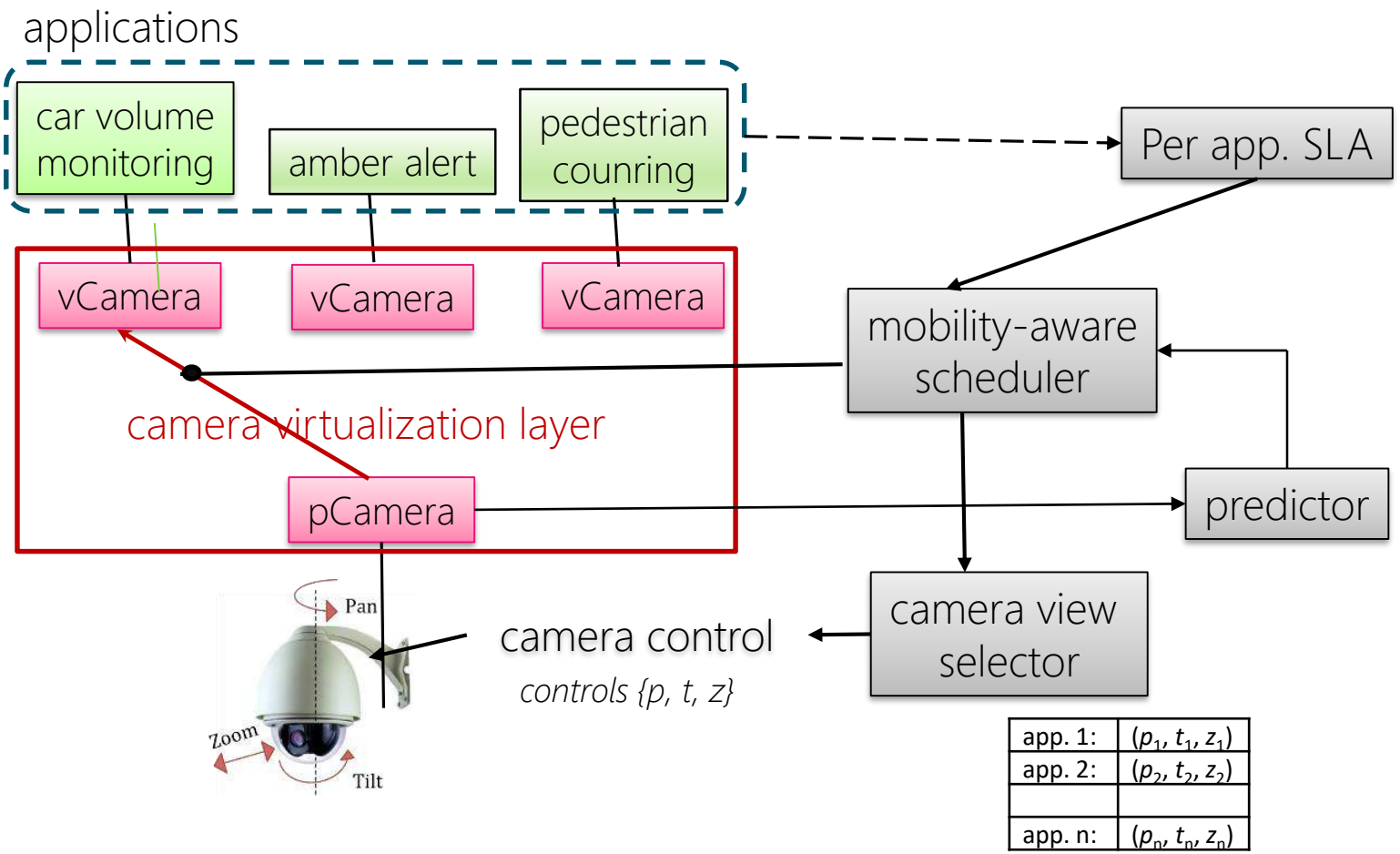
break one-to-one binding between camera & application



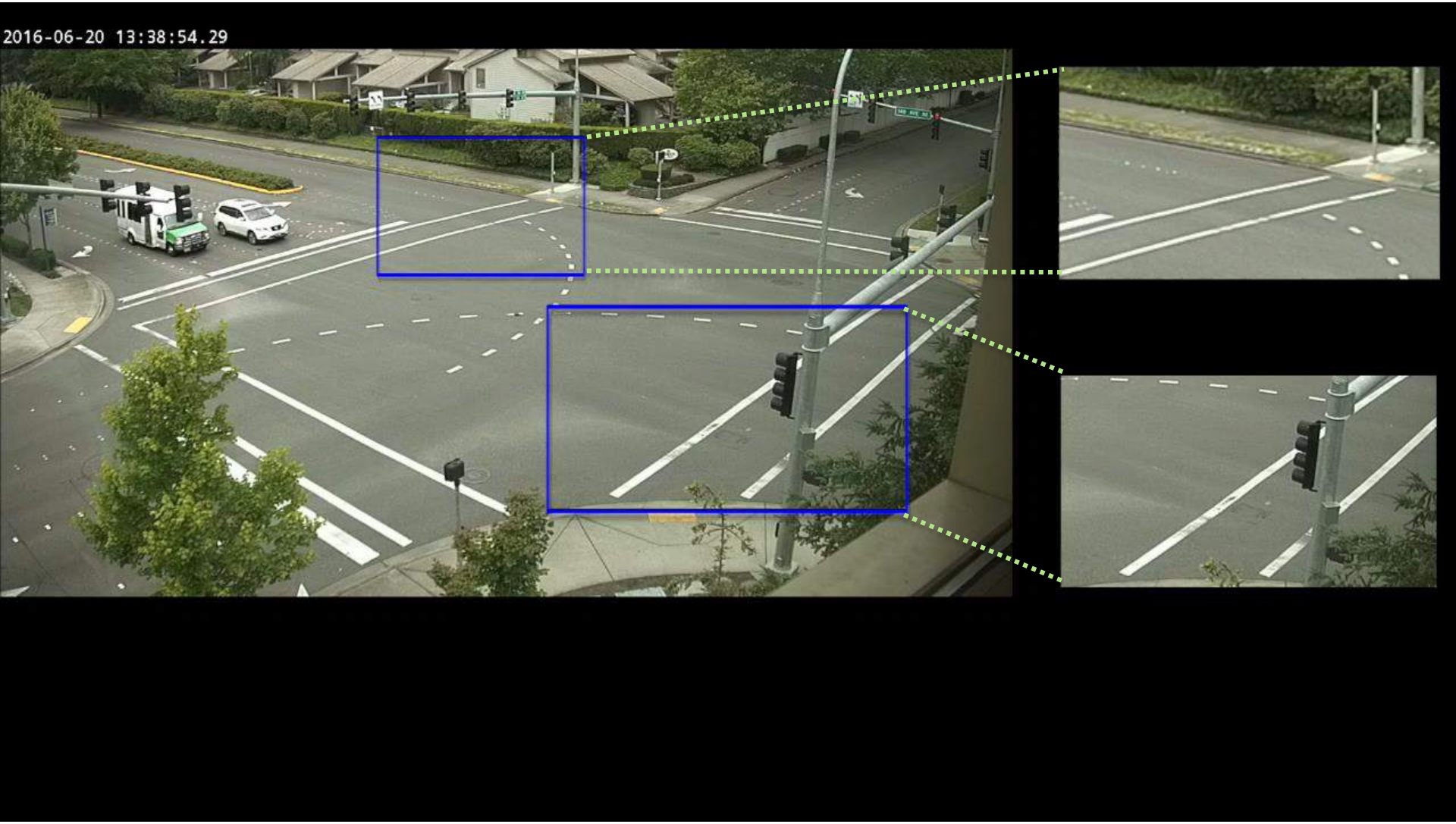
Microsoft

camera management system

Shubham



Microsoft


2016-06-20 13:38:54.29

MCS 2013

Vision: Cloud-Powered Sight for All

Showing the Cloud What You See

Paramvir Bahl Matthai Philipose Lin Zhong

best paper award

MobiSys 2013

Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision

Robert LiKamWa[†,‡], Bodhi Priyantha[‡], Matthai Philipose[‡], Lin Zhong[†,‡], and Paramvir Bahl[‡]

SenSys 2014

Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices

Tiffany Yu-Han Chen Lenin Ravindranath Shuo Deng

sigmobile highlights

SEC 2017

ParkMaster: An in–vehicle, edge–based video analytics service for detecting open parking spaces in urban environments

Giulio Grassi Paramvir Bahl

MobiCom

The Design and Implementation of a Wireless Video Surveillance System

NSDI 2017

Live Video Analytics at Scale with Approximation and Delay-Tolera

Haoyu Zhang*†, Ganesh Ananthanarayanan*, Peter Bodik*, Matthai Philipose*

IPSNI 2017

Panoptes: Servicing Multiple Applications Simultaneously Steerable Cameras

Shubham Jain
WINLAB, Rutgers University
shubhamj@winlab.rutgers.edu

Viet Nguyen
WINLAB, Rutgers University
viet.nguyen@rutgers.edu

Marco Gruteser
WINLAB, Rutgers University
gruteser@winlab.rutgers.edu

Paramvir Bahl
Microsoft Research
bahl@microsoft.com

ABSTRACT

Steerable surveillance cameras offer a unique opportunity to support multiple vision applications simultaneously. However, state-of-art camera systems do not support this as they are often limited to one application per camera. We believe that we should break the one-to-one binding between the steerable camera and the application. By doing this we can quickly move the camera to a new view needed to support a different vision application. When done well, the scheduling algorithm can support a larger number of applications over an existing network of surveillance cameras. With this in mind we developed *Panoptes*, a technique that virtualizes a camera view and presents a different fixed view to different applications. A scheduler uses camera controls to move the camera appropriately providing the expected view for each application in a timely manner, minimizing the impact on application performance.

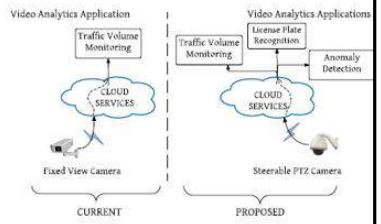


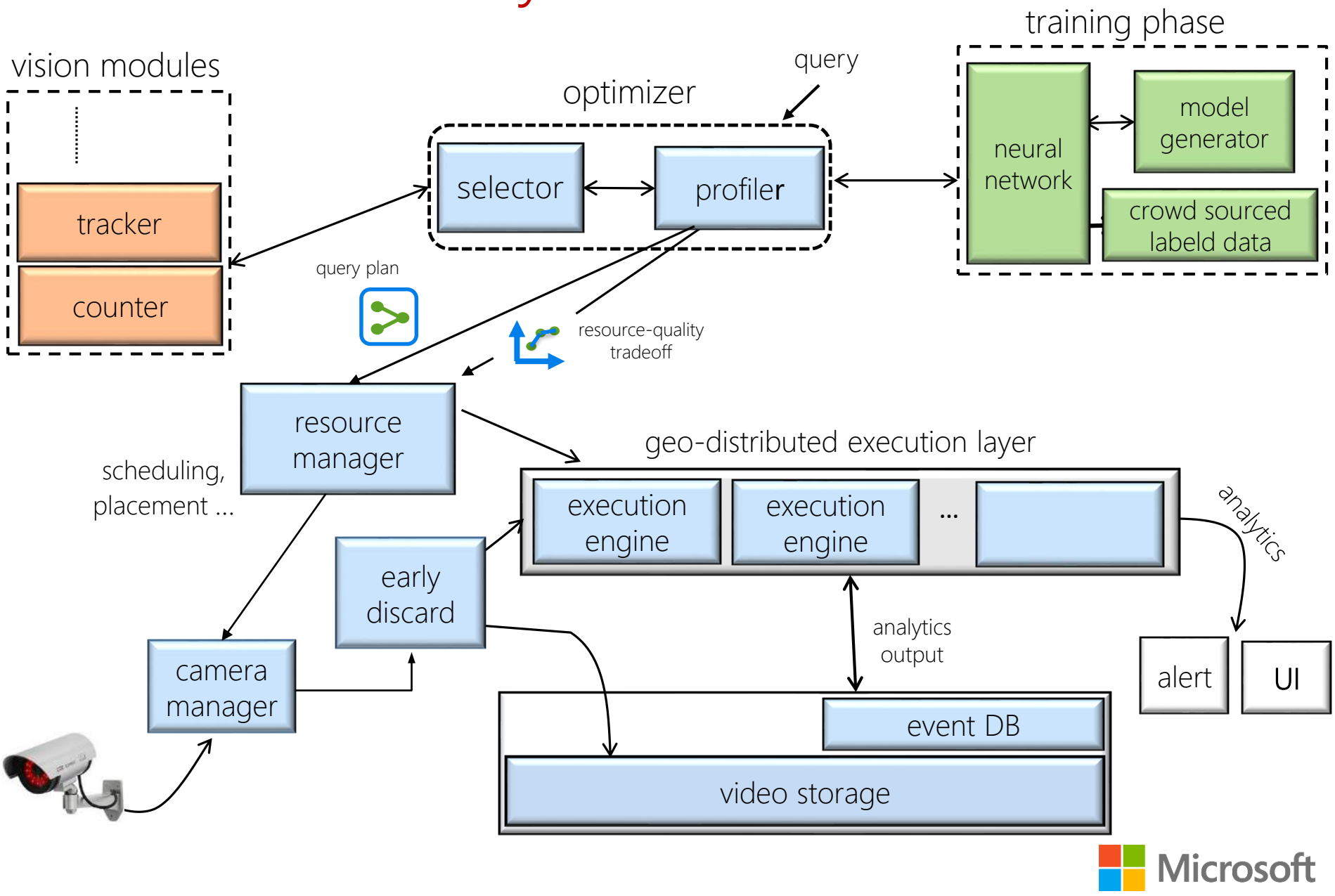
Figure 1: Cloud analytics platform for cameras.

Microsoft

deployment

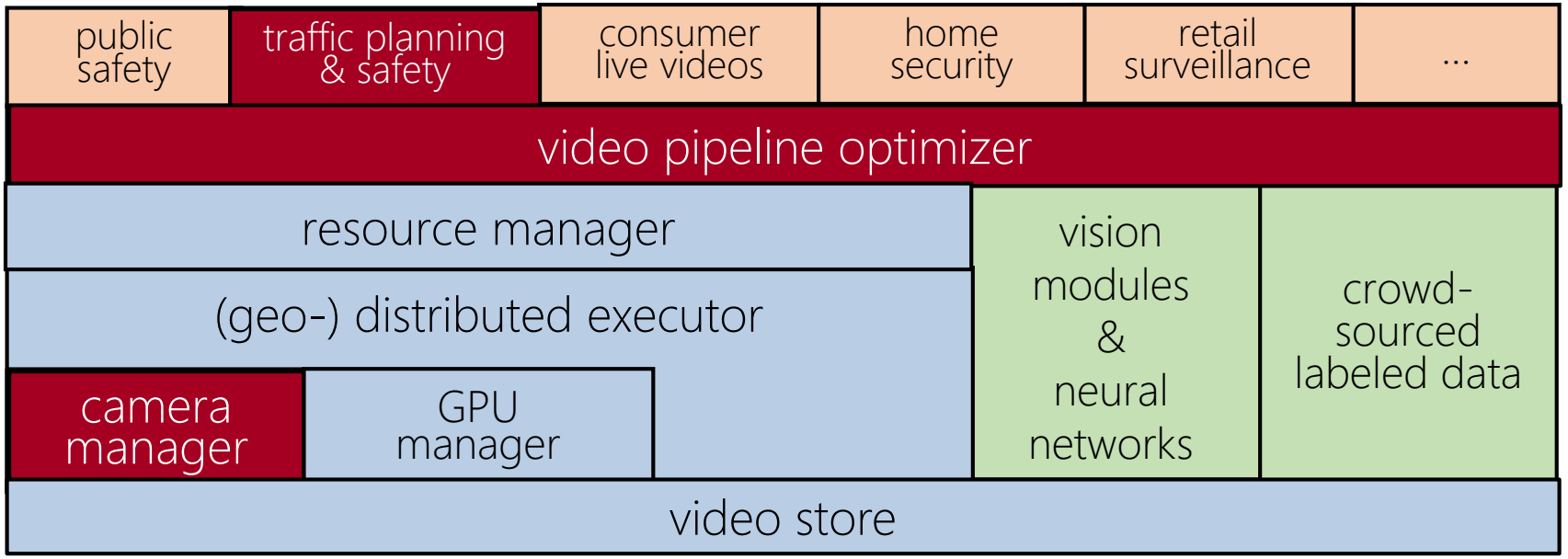
accuracy • latency • bandwidth • cost

the system we built



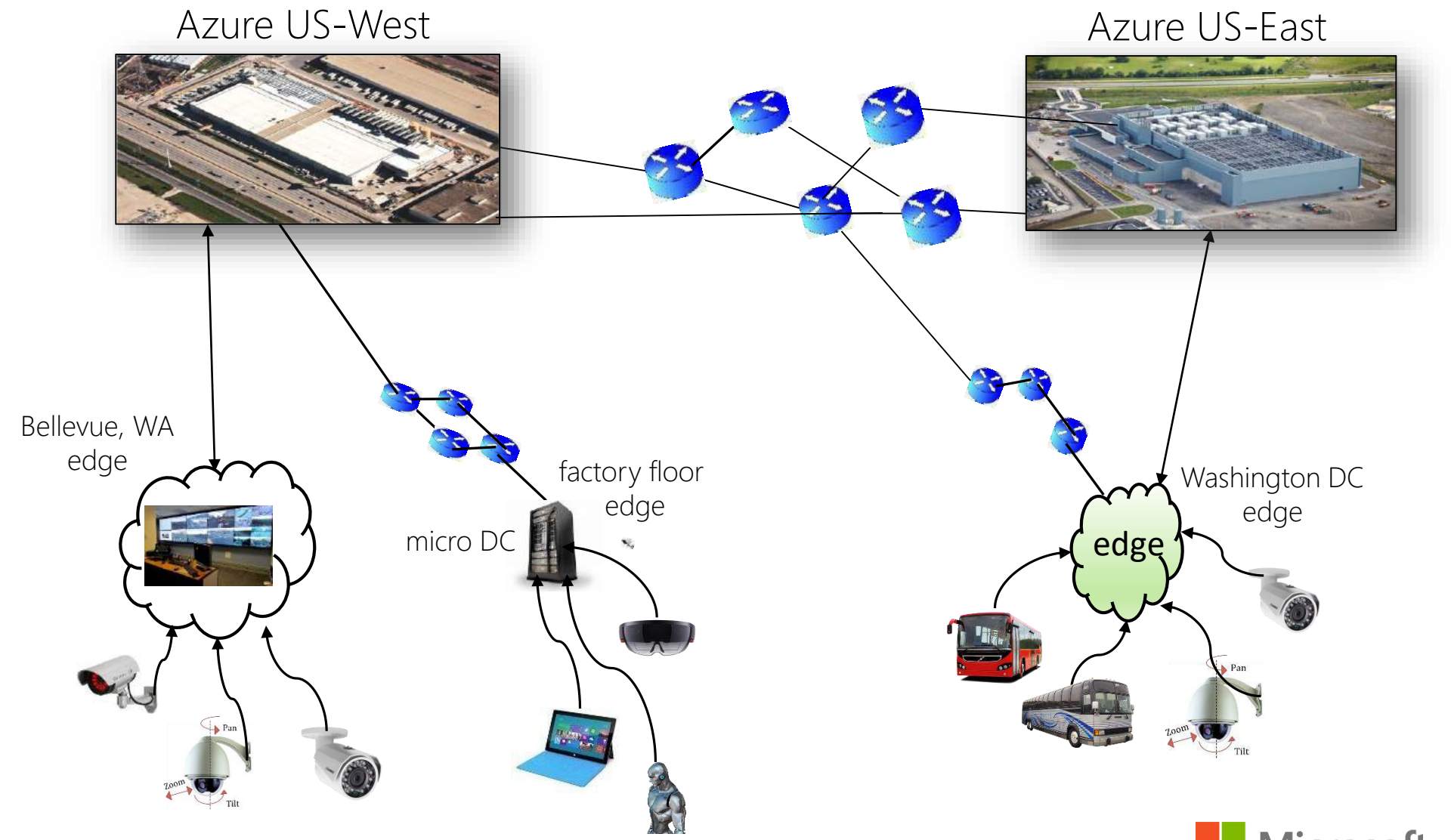
Microsoft

the stack we built: MSR's Rocket



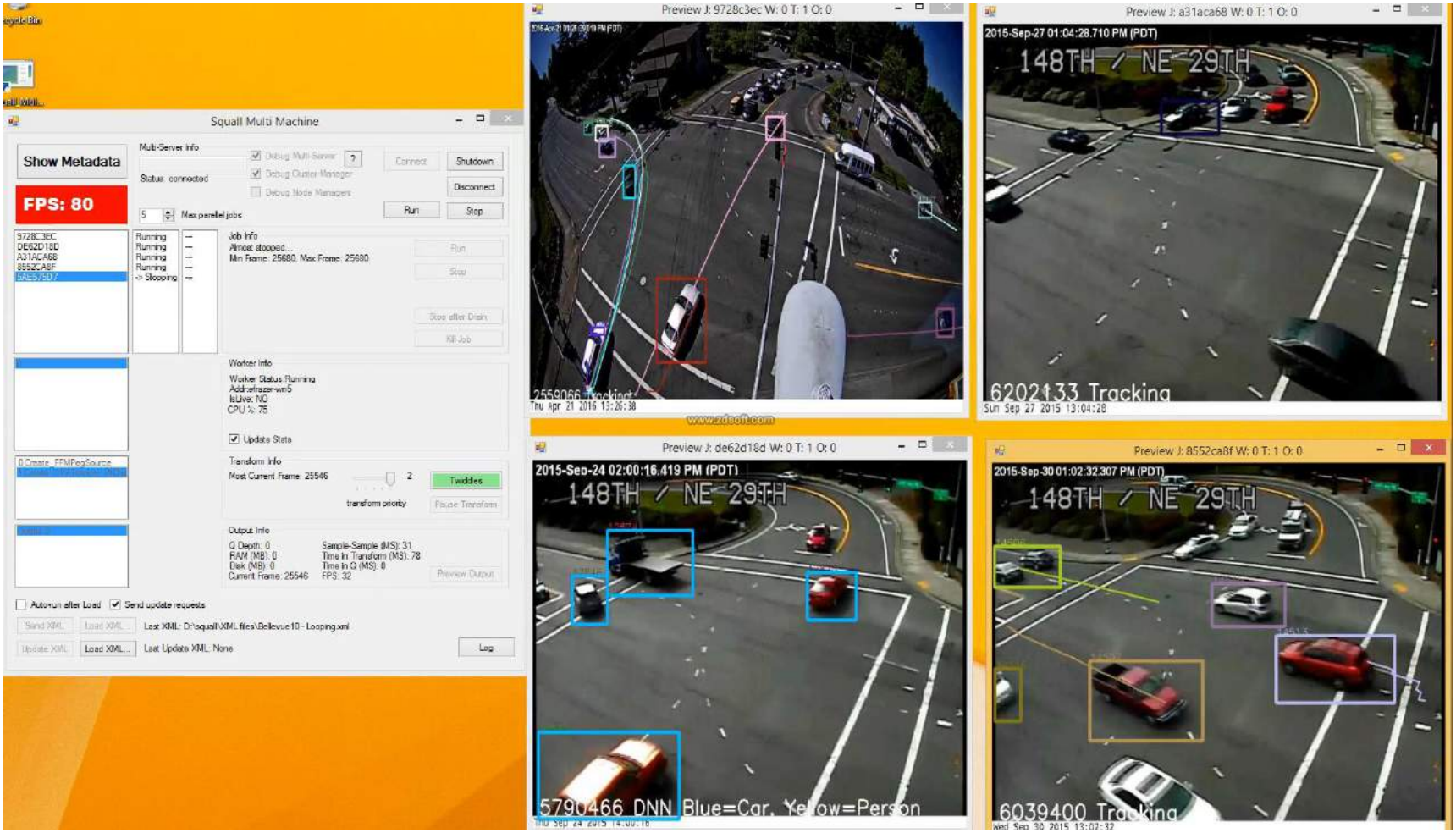
Microsoft

deployment: hybrid edge-cloud architecture



Microsoft

multi-camera implementation in Bellevue



Microsoft

live dashboard

http://vavz.azurewebsites.net/

Microsoft Traffic Analytics

Dashboard

BellevueWay ... Sep 04 11:19am - 1...

116th NE1... Sep 04 11:19am - 12:1...

150th SE3... Sep 04 11:19am - 12:1...

150th New... Sep 04 11:19am - 12:1...

Busiest Intersections

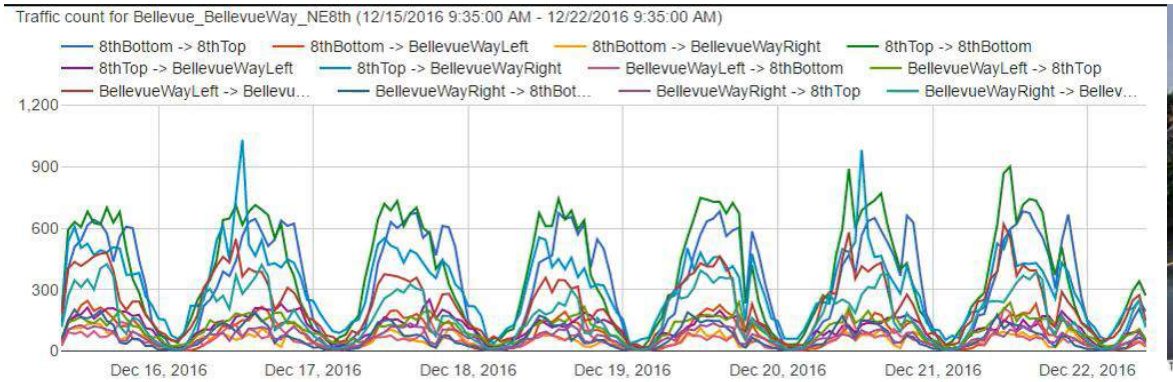
Location	Last Hr Total	Week ADT
BellevueWay NE8th	2165	35867
150th Eastgate	851	23106
150th SE38th	846	28305
116th NE12th	752	17419
150th Newport	693	13803

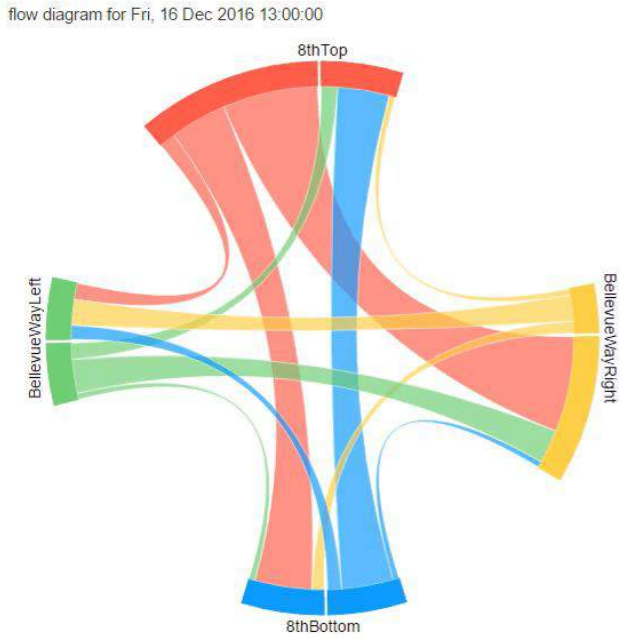
Alerts

Location	Last Hr Total	Avg Hr Total
116th NE12th	752	1631
150th Eastgate	851	1987
150th SE38th	846	1895

Microsoft

direction counting accuracy





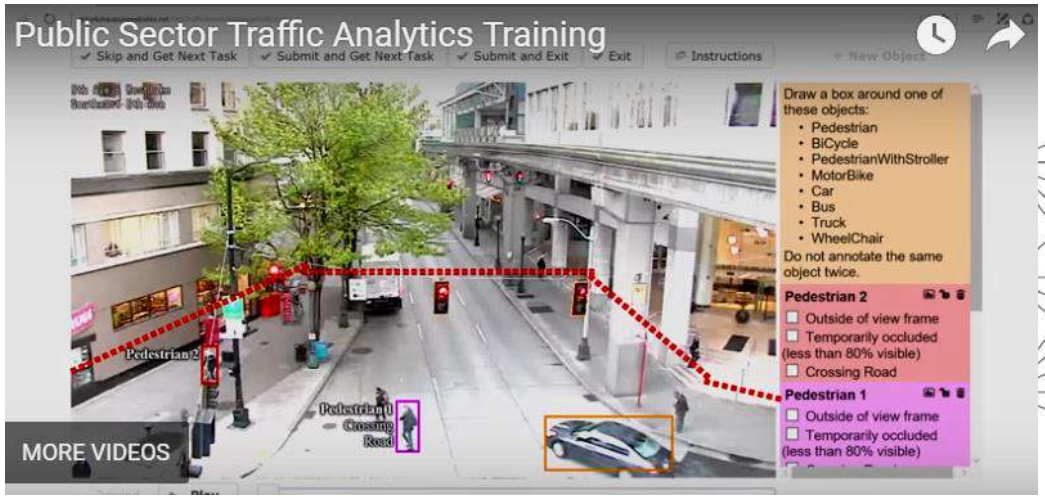
12 directions (lane-wise counts) occlusions due to 3D → 2D projection on lanes

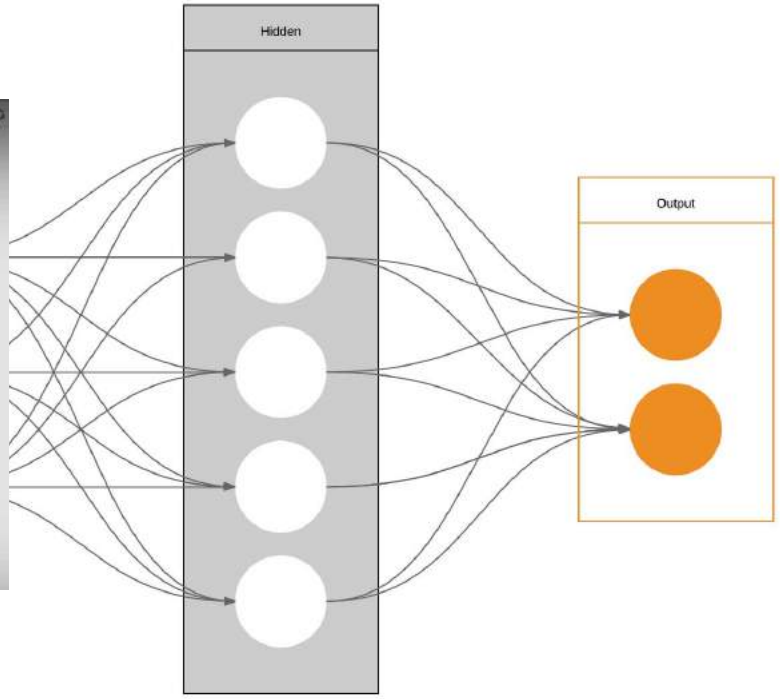
95% count accuracy compared to crowdsourced ground truth

Microsoft

training neural networks

labeled data



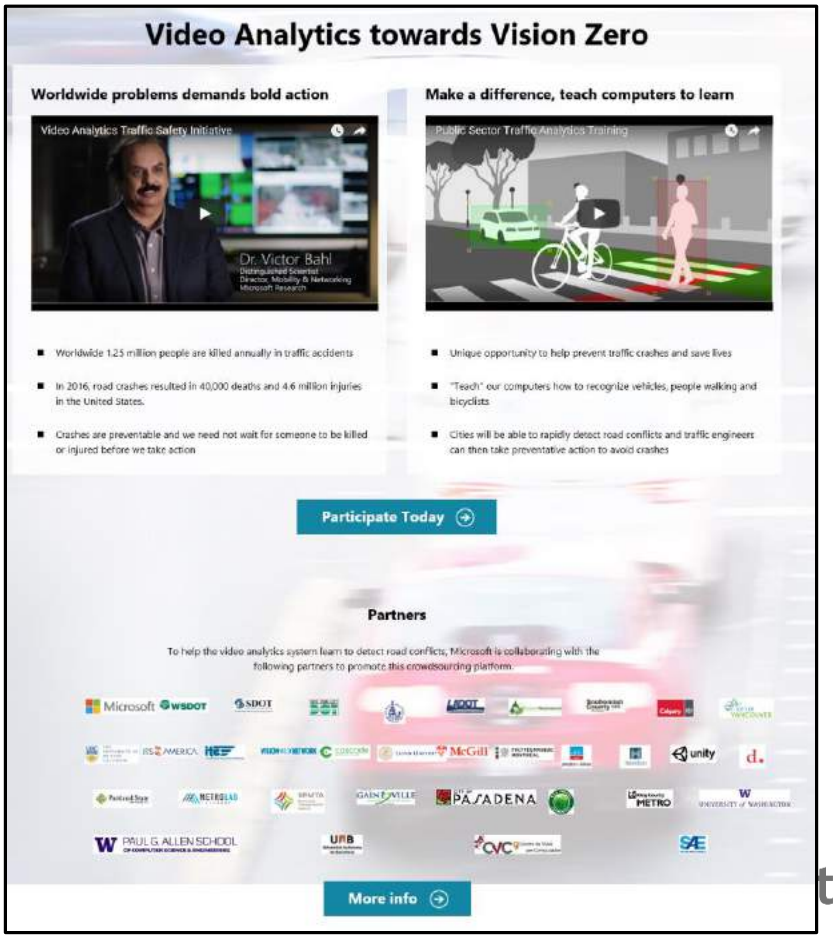


Microsoft

national initiative to train NN

(launched July 1, 2017)

http://www.ite.org/visionzero/videoanalytics/



can we solve all problems?
(can humans do better?)

actively reducing accidents

making self-driving cars safer

(10 million self-driving cars by 2020 – Forbes, March 2017)

network 2

network 1

1080p - 4k @ 30fps

Azure

>2 Mbps LTE

city-level network
inexpensive
scalable across city
moderately reliable

> 40 Mbps Ethernet
hi-res video

edge node

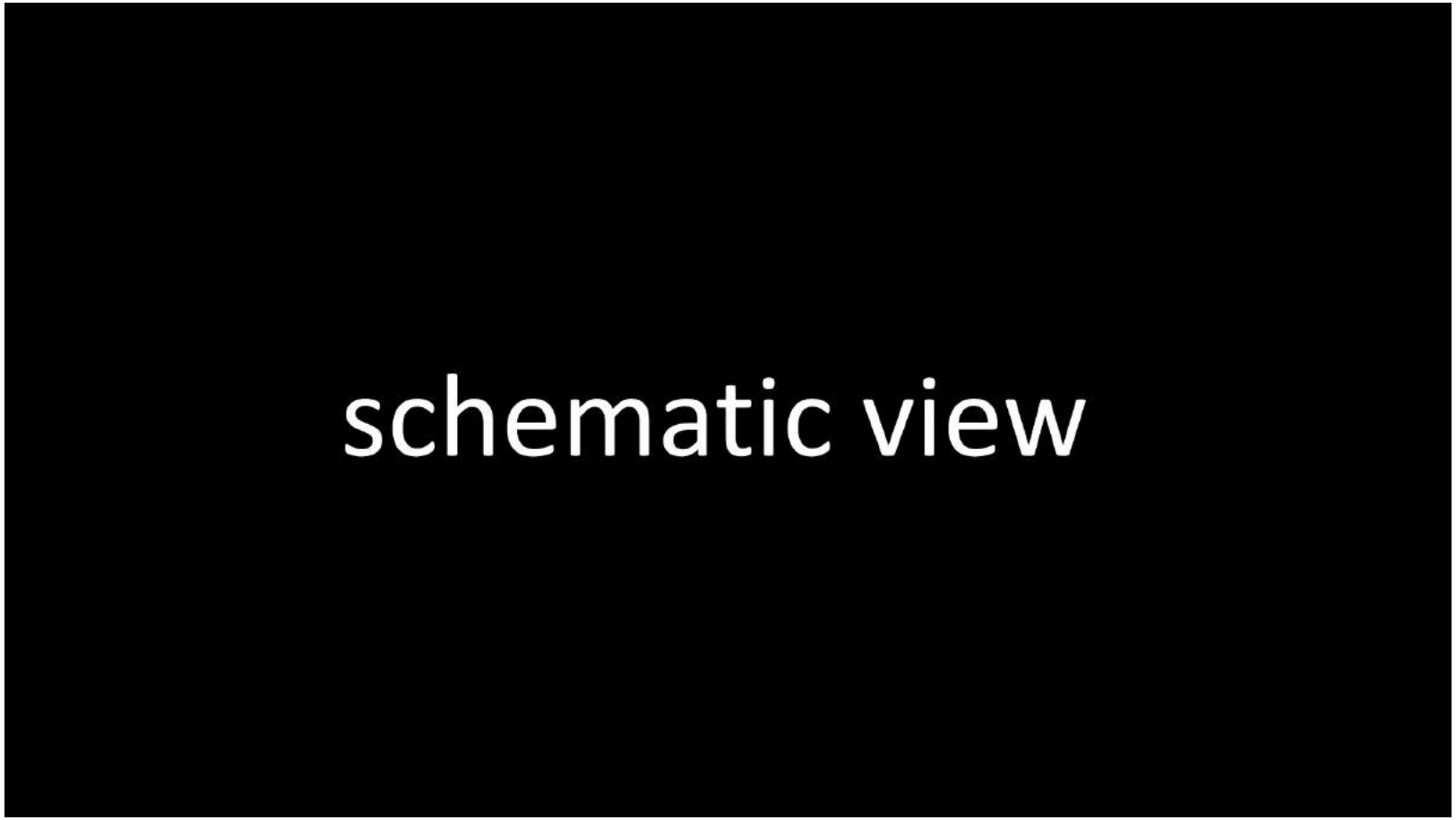
GPU @ >500 GFLOPS
+ CPU + HW codec

10 kbps DSRC
control messages

street-level network
reliable,
low-latency,
autonomous,
locally scalable

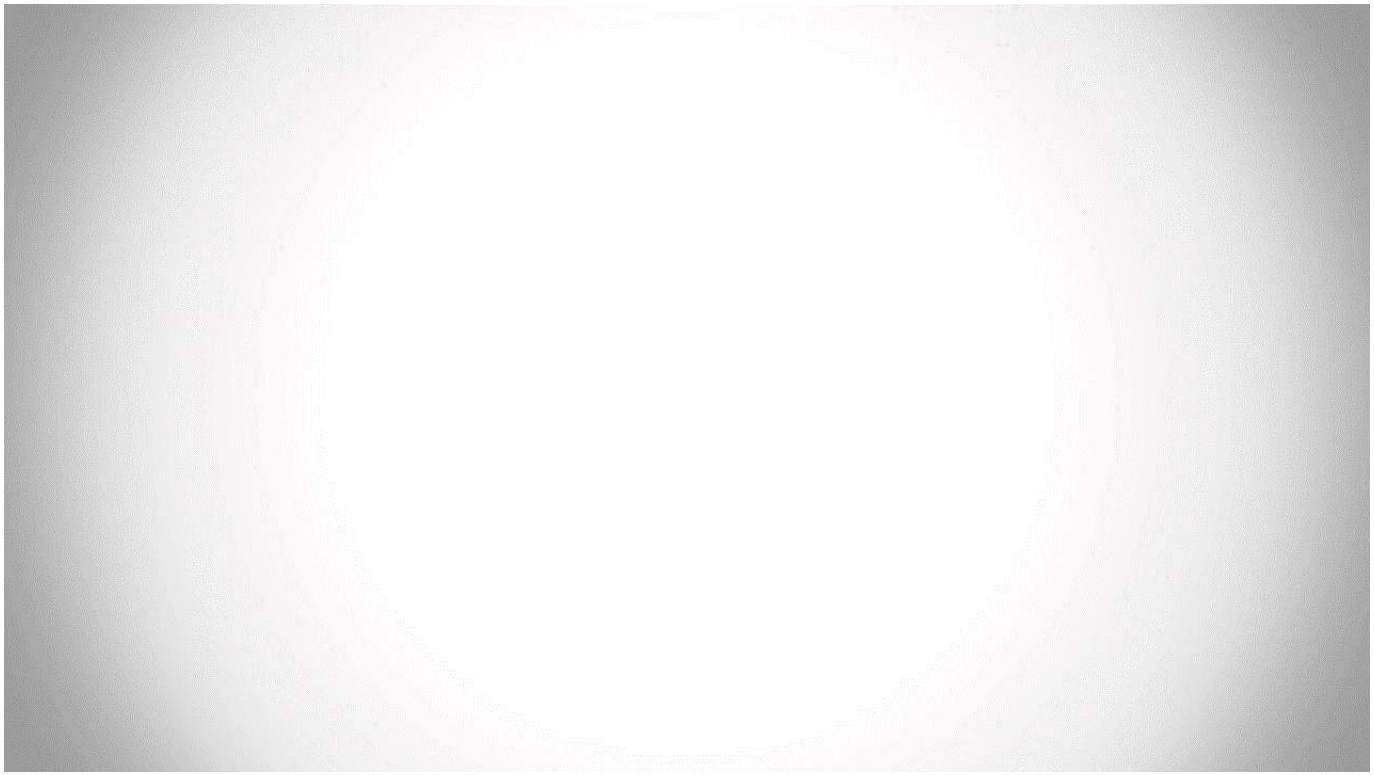
Microsoft

live demonstration in Hannover Messe



Microsoft

some nice memories



Microsoft

mayor's challenge award to Bellevue

U.S. Department of Transportation

MAYORS' CHALLENGE

GATHER DATA

Winner!
Bellevue, Washington

Bellevue, WA, pursued a range of data collection activities during the Mayors' Challenge to identify barriers to bicycling and walking, prioritize improvements, and guide investments. In February 2015, the Bellevue City Council introduced the Pedestrian and Bicycle Implementation Initiative (PBII) to improve safety for people of all ages and abilities who walk and bike in Bellevue. Using data collected from online sources, key-pad polling at public meetings, automated bicycle and pedestrian counters, and traffic camera videos, the PBII team identified barriers to walking and bicycling and developed a $6.8M Bicycle Rapid Implementation Program (BRIP) budget proposal to guide citywide investments through 2019. The BRIP aims to expand the city's bicycle network from 42 miles to more than 70 miles of conventional bike lanes, separated lanes or off-street paths, and to complete four continuous, cross-city bicycle corridors.

Demonstrated Successes

Innovative Data Collection Techniques Gather Real-Time and Long-Term Data with Public Input

Throughout the PBII process, Bellevue has emphasized understanding long-term trends and gathering feedback from people who walk and bike. Bellevue's PBII team:

- Conducted a longitudinal assessment from 2006–2015 of non-motorized collisions using the USDOT's Pedestrian and Bicycle Crash Analysis Tool (PBCAT) system;
- Gathered input using key-pad polling and comment cards at 20 public meetings and an open house that attracted 140 attendees; and
- Used online surveys to solicit public input at two stages in the BRIP development process;
  - Over 700 people placed more than 1,600 points in the first online map to identify locations that they felt were unsafe for walking and bicycling;
  - Over 120 people submitted more than 400 comments on conceptual designs for 52 proposed projects to make the pedestrian and bicycle systems safer.

Pedestrian and Bicycle project manager Franz Loewenherz (foreground) and Councilmember Lynne Robinson (center) lead a policy ride with local bicycle advocates in Downtown Bellevue.

Microsoft

Video Analytics analyzes traffic camera video footage and uses near-miss collisions to predict where future crashes are likely to occur. Traffic engineers could then take corrective action to prevent them. File photo

Bellevue video analytics project receives safety award

Fri Aug 4th, 2017 3:44pm • BUSINESS

U.S. Department of Transportation

Microsoft

Computer

COVER FEATURE **EDGE COMPUTING**

Real-Time Video Analytics: The Killer App for Edge Computing

Ganesh Ananthanarayanan, Paramvir Bahl, Peter Bodík, Krishna Chintalapudi, Matthai Philipose, Lenin Ravindranath, and Sudipta Sinha, Microsoft Research

Video analytics will drive a wide range of applications with great potential to impact society. A geographically distributed architecture of public clouds and edges that extend down to the cameras is the only feasible approach to meeting the strict real-time requirements of large-scale live video analytics.

According to a 2015 report by the Information Handling Services on the installed base for video surveillance equipment, there is a camera installed for every 29 people on the planet, with mature markets having a camera for every 8 people. The report predicts that the number of cameras will grow by 20 percent year over year for the next 5 years. Video analytics from these cameras are used for traffic control, surveillance, and security in both public and private venues, as well as consumer applications including digital assistants for real-time decisions.

We propose that a geographically distributed architecture of public clouds, private clusters, and edges that extend down to the cameras is the only approach that can meet the strict real-time requirements of large-scale video analytics, which must address latency, bandwidth, and provisioning challenges.

First, applications require very low latency when processing video because the output of the analytics is used to interact with humans (such as in augmented reality scenarios) or to actuate some other system (such as traffic lights). Second, high-definition video requires large bandwidth—5 Mbps or even 25 Mbps for 4K video—and streaming a large number of video feeds directly to the cloud might not be feasible. When cameras are connected wirelessly, such as inside a car, the available uplink bandwidth is very limited. Finally, using compute capacities available on the camera itself allows for correspondingly lower provisioning (or usage) in the cloud. This also means that less interesting parts of the video can quickly be filtered out, for example, using motion-detection techniques, which dramatically reduces the bandwidth that needs to be provisioned.

Aside from low latency and efficient bandwidth usage, another major consideration for continuous video analytics is video processing's high compute cost. Because of high data volume, compute demands, and latency requirements, cameras are the most challenging "things" in the Internet of Things. Thus, large-scale video analytics could well be edge computing's "killer app." Tapping into the

0018-9162/17/$33.00 © 2017 IEEE

IEEE

IET The Institution of Engineering and Technology

PARTNERSHIPS

Video Analytics Towards Vision Zero

By Franz Loewenherz, Victor Bahl, Ph.D., and Yinhai Wang, Ph.D.

For young people below the age of 35, motor vehicle crashes are the leading cause of death in the United States. In 2015, collisions resulted in 35,092 deaths and 2.4 million injuries. More than 1,100 children under the age of 15 were killed. The 7.2 percent increase in traffic fatalities from 2014 to 2015 represents the greatest percentage increase in nearly 50 years.[1] Yet despite the massive death toll, work to prevent traffic fatalities has been woefully lacking.

Many governmental agencies continue to rely on traditional traffic safety approaches. They intervene only after enough police crash reports are filed to trigger a High Crash Corridor designation. This reactive approach to preventing crash recurrence has well-documented limitations:

- At most locations, the number of crashes is very small and subject to chance variations;
- Not all crashes are reported and the level of reporting is uneven regarding the type of road users involved, the exact location, and the severity of injuries;
- Numerous "close calls" go undocumented; and
- Many years of crash data are typically required to develop an understanding of the situation.[2]

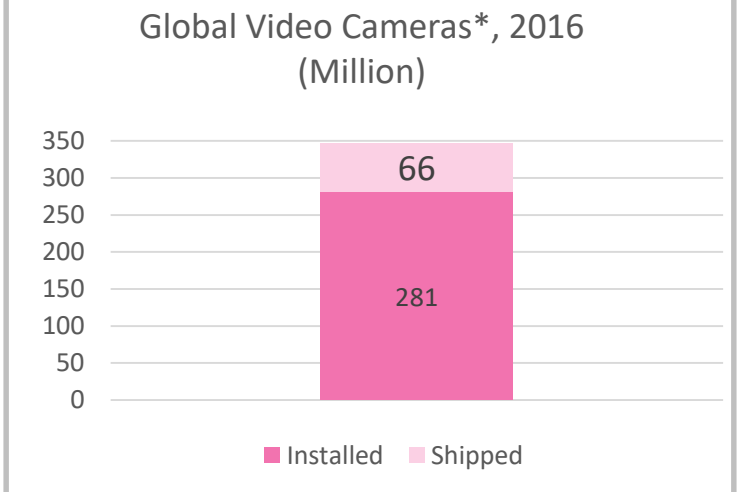
Given these trends, and the crash analysis tools presently employed, how will jurisdictions achieve what all of us want: zero fatalities and serious injuries on our roadways? That's the goal of Vision Zero, [illegible] what is one of the leading causes of death worldwide.[3] It calls on government agencies to be proactive, identify risks, and take steps to prevent injuries on our roadways. Vision Zero encourages us to imagine a future in which we do not need to wait for crashes to occur in order to prevent others from happening.[4]

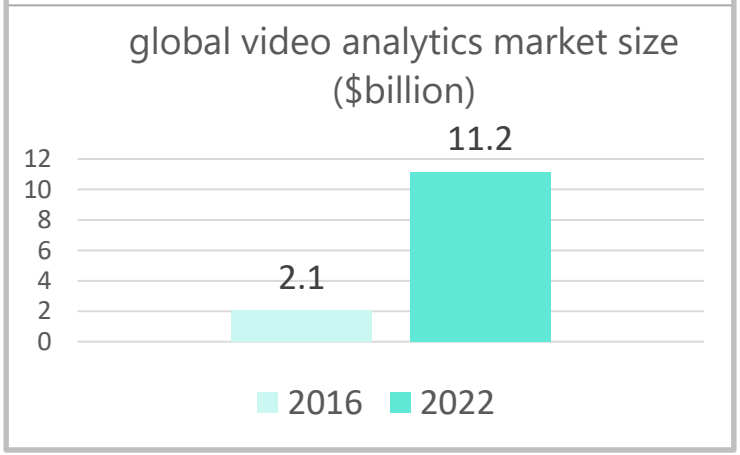
Solutions for a Safer World

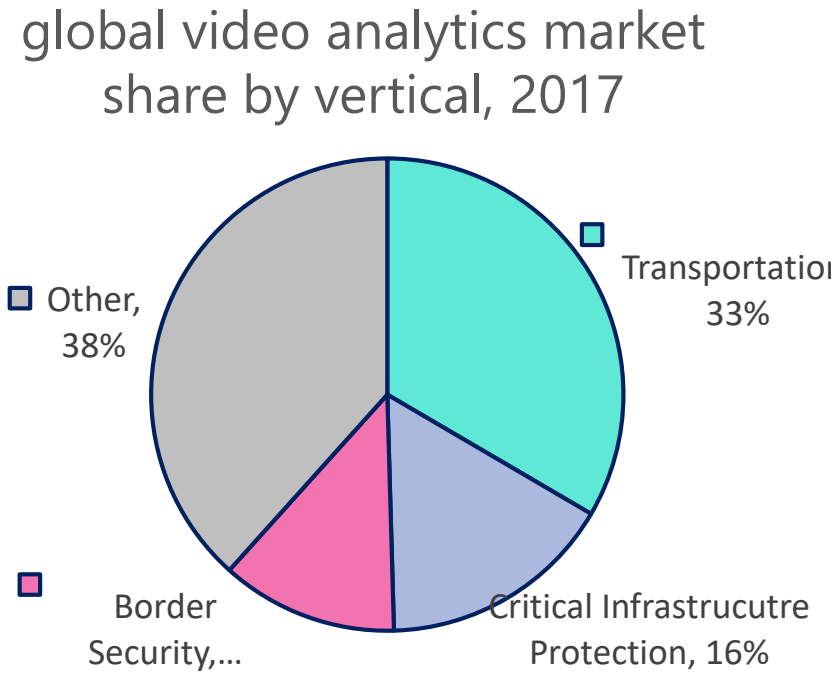
Although traffic collisions can happen anywhere, there are often early warning signals in the form of conflicts or near-collision events at specific locations. These are recurring instances where a car abruptly stops because a bicycle veered in front of it, a pedestrian steps into the path of a bicyclist, or one bicyclist or car passes by another or a static object at very close spacing. These surrogate warning indicators – observable non-crash traffic conflict events – provide insight into when, where, and why crashes are most likely to occur. Understanding the root causes for near-collision events could enable local governments to take proactive, [illegible]shes.[5]

opportunity for AI in video analytics

market opportunity







*59% of installed cameras in 2016 are IP cameras
Source: Markets&Markets, IHS market

Microsoft

… but then something interesting happened …


"it's not the mobile devices but all the other things that are out at the edge that are truly going to transform cloud computing and put an end to what we know as the cloud.
Peter Levine
Andreessen Horowitz
December 16, 2016


"we're moving from what is today's mobile-first, cloud-first world to a new world that is going to be made up of an intelligent cloud and an intelligent edge"
- Satya Nadella
CEO Microsoft
Build, May 10, 2017

Gartner.

Maverick* Research: The Edge Will Eat the Cloud

Published: 22 September 2017 **ID:** G00338633

Analyst(s): *Thomas J. Bittman*

Summary

The growth of the Internet of Things and the upcoming trend toward more immersive and interactive user interfaces will flip the center of gravity of data production and computing away from central data centers and out to the edge. (Maverick research exposes unconventional thinking and advice.)

More on This Topic

This is part of an in-depth collection of research. See the collection:

SERIES OVERVIEW

Maverick* Insights Drive Creative Destruction of Business as Usual: A Gartner Trend Insight Report (https://www.gartner.com/document/code/342945?ref=grbody&refval=3806165)

Overview

Specific Maverick Caution

This research contradicts prevailing views on the future of cloud computing, the topology of computing architectures and the nature of applications as we move toward digital business. Instead of continued growth of mega data centers, compute and storage will move toward the edge, due to the Internet of Things and new user/machine interfaces.

"forty percent of large enterprises will be integrating edge computing principles into their 2021 projects, up from less than 1% in 2017"

2017 the year edge computing took off

TechCrunch

Microsoft is extending Azure IoT to the edge of the network

Posted May 10, 2017 by Frederic Lardinois (@frederic)

Forbes

JUN 7, 2017 @ 01:54 PM 9,495

Amazon Makes Foray Into Edge Computing With AWS Greengrass

Janakiram MSV, CONTRIBUTOR

I cover Cloud Computing, Machine Learning, and Internet of Things FULL BIO

Opinions expressed by Forbes Contributors are their own.

CRN

Dell Technologies Launches A New IoT Division And Partner Program Amid 'Great Boom' In Edge Computing

by Kyle Alspach on October 10, 2017, 11:26 am EDT

TechRepublic

AT&T wants to reinvent the cloud with low latency edge computing over 5G

With new investments in edge computing, AT&T could set the stage for business investments in IoT, VR, robotics, and more.

By Conner Forrest | July 18, 2017, 6:20 AM PST

MOBILE EUROPE

Deutsche Telekom targets 5G applications with edge computing testbed

20 OCTOBER 2017

IT WORLD

GE adds edge analytics, AI capabilities to Predix industrial IoT suite

At its annual Minds + Machines conference, GE Data is unveiling a slew of extensions to its Predix Edge PaaS

By Marc Ferranti

Editor in Chief, IDG News Service | OCTOBER 24, 2017

Microsoft

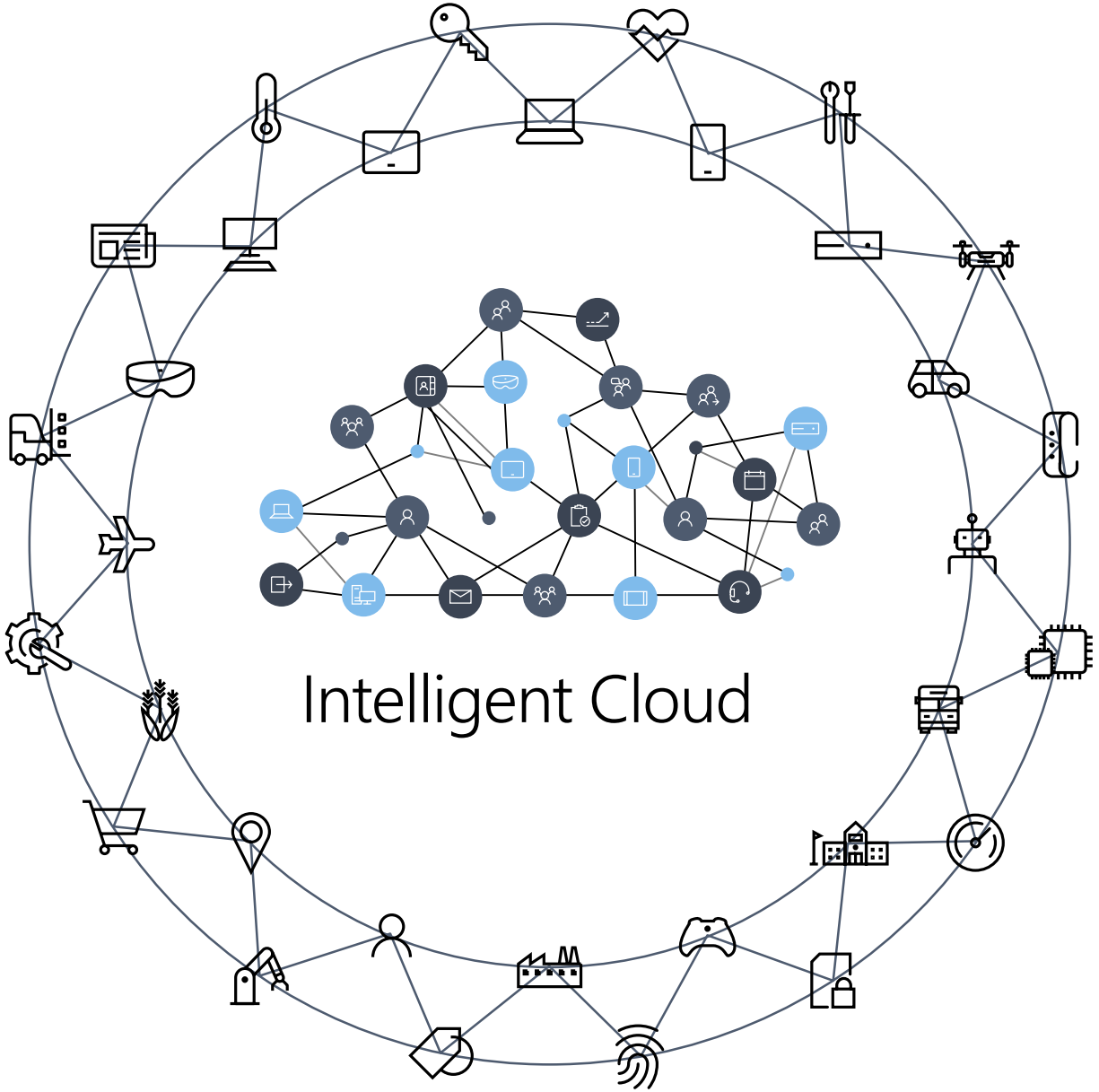
Microsoft will invest $5 billion in IoT. Here's why.

Apr 4, 2018 | Julia White, CVP Microsoft Azure

Advancing the Opportunity in IoT and Intelligent Edge

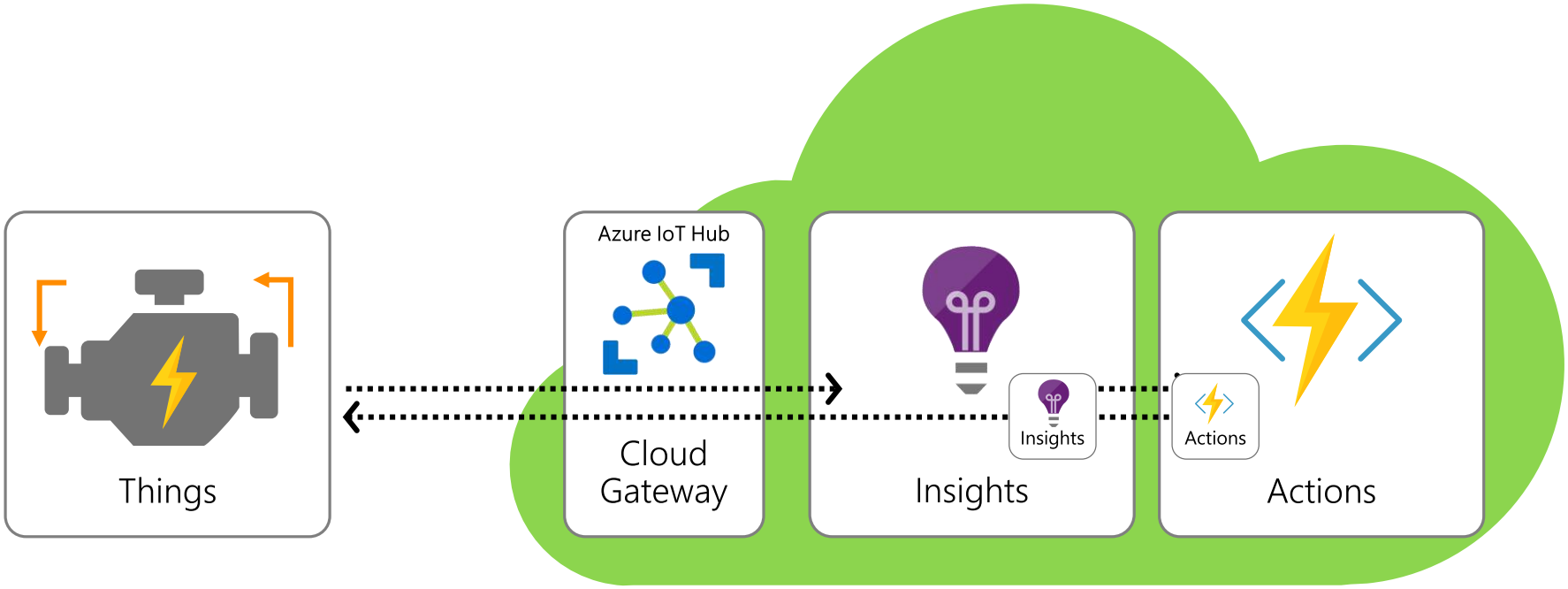
We've been invested in IoT before the term was coined, when enterprises had these endpoints in their factories, buildings and other devices that were totally "dark." Today, we're planning to dedicate even more resources to research and innovation in IoT and what is ultimately evolving to be the new intelligent edge.

Microsoft



Microsoft

Azure's perspective on IoT App pattern + Edge



Azure IoT edge

Cloud services at the edge

Azure ML, Azure Stream Analytics, Azure Functions, custom

Manage from the cloud

Devices and services from Azure Portal

Flexible connectivity

Intermittent, low, or no connectivity

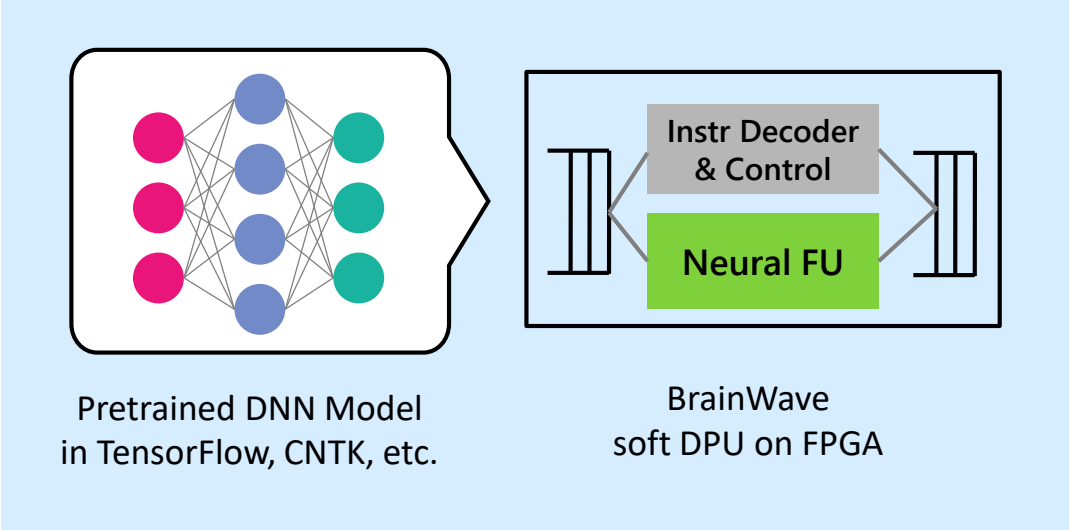
Reduced latency and cost

Bring compute to the data, reduced bandwidth cost

project Brainwave @ the edge

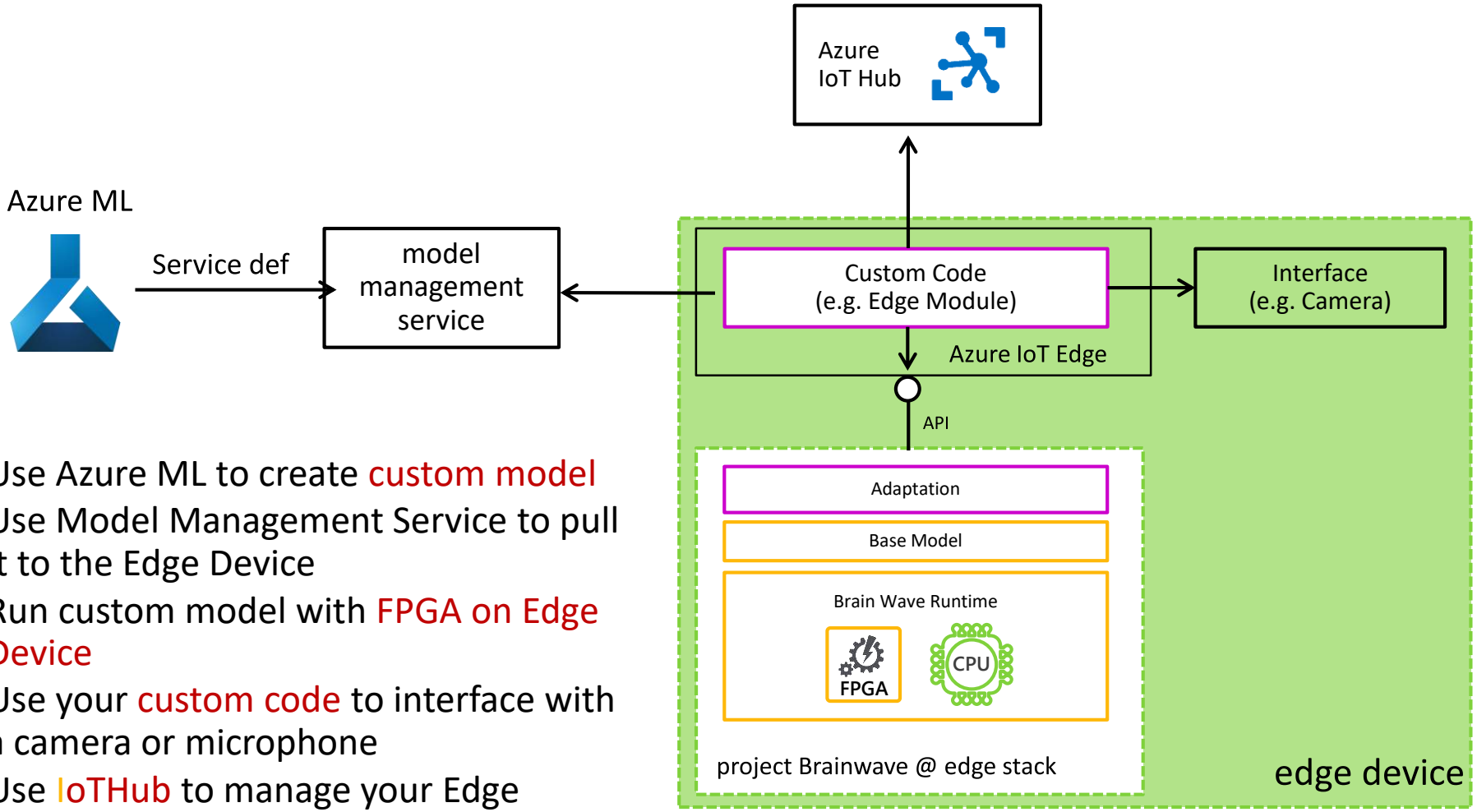
an accelerated FPGA powered AI Platform:

- Fast: Ultra-low latency, high-throughput serving DNN models at low batch sizes
- Flexible: Future proof, adaptable to fast-moving AI space & evolving model types
- Friendly: Turnkey deployment of TensorFlow/CNTK/Caffe/etc.



Microsoft

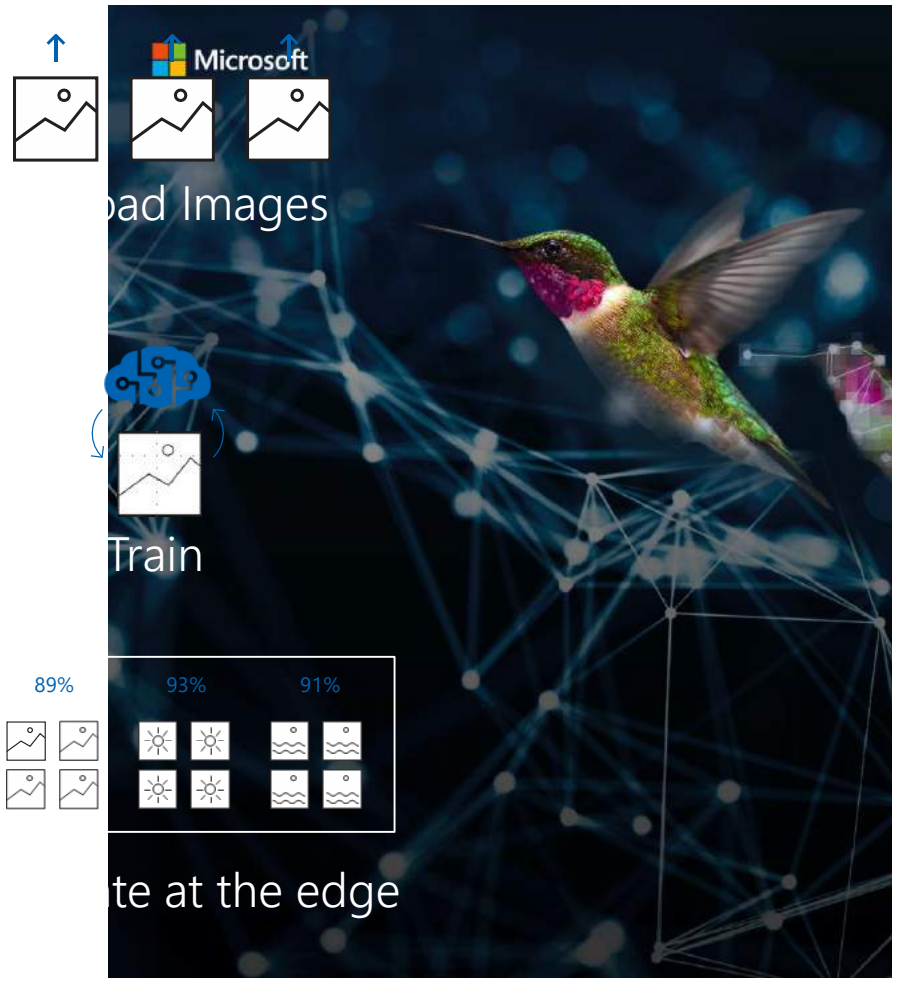
deploying and running a model



1. Use Azure ML to create custom model
2. Use Model Management Service to pull it to the Edge Device
3. Run custom model with FPGA on Edge Device
4. Use your custom code to interface with a camera or microphone
5. Use IoTHub to manage your Edge Module & data streams

Microsoft

AI at the “cutting edge



- defect detection
- video surveillance
- example

Microsoft

what is the edge?

Sensor Tier

Constrained Tier

Interactive Tier

Industrial Tier

Gateway Tier

GPU/DSP FPGA

Accelerated Tier

LOW POWER CAPABILITIES

HIGH POWER CAPABILITIES

Azure IoT Edge hardware requirements

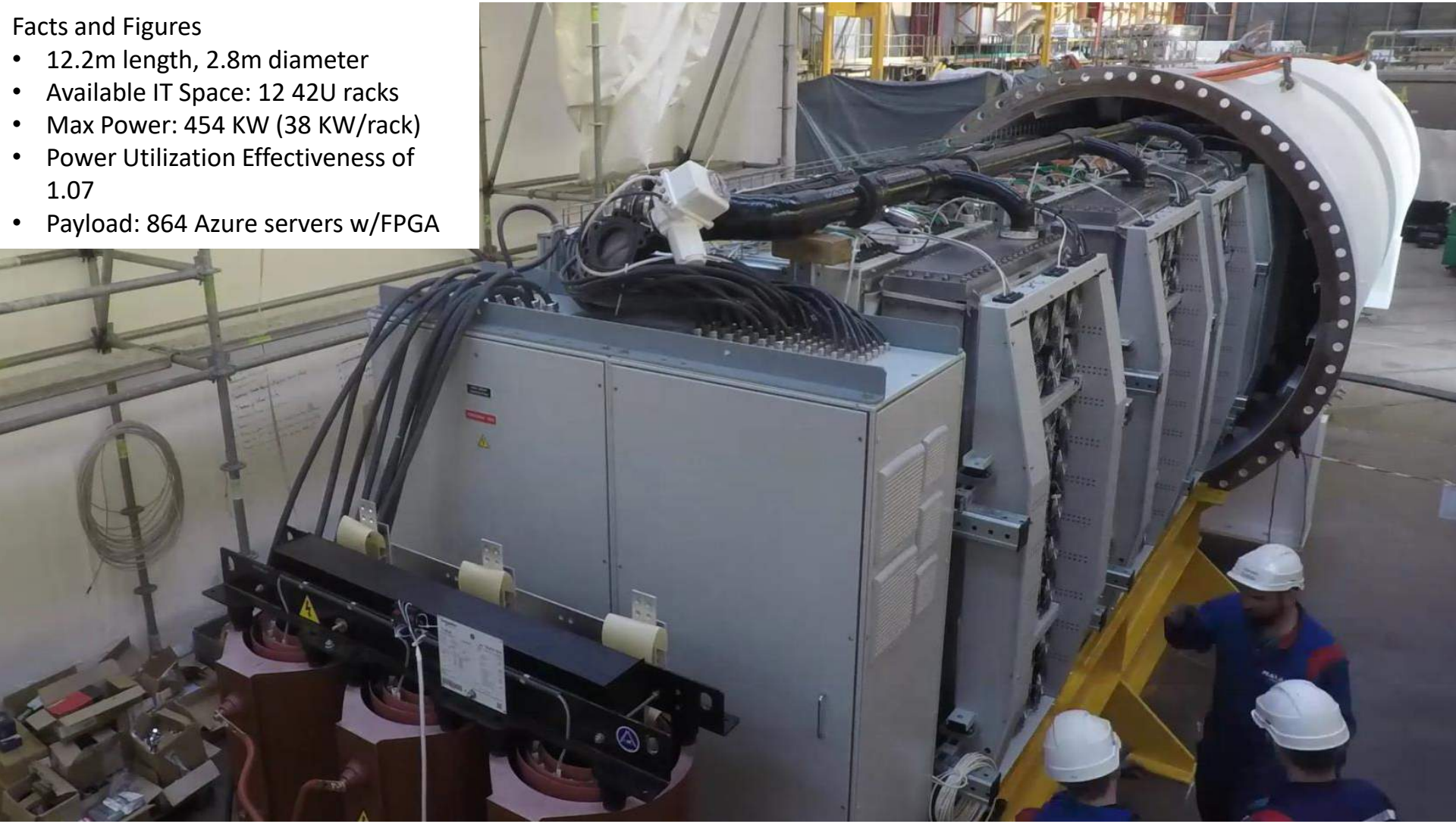
- Rich OS - Windows or Linux
- flexible HW – ARM or x64
- Moby-compatible container runtime
- hardware based security – HSM or Enclave
- hardware sizing depends on workload

Microsoft

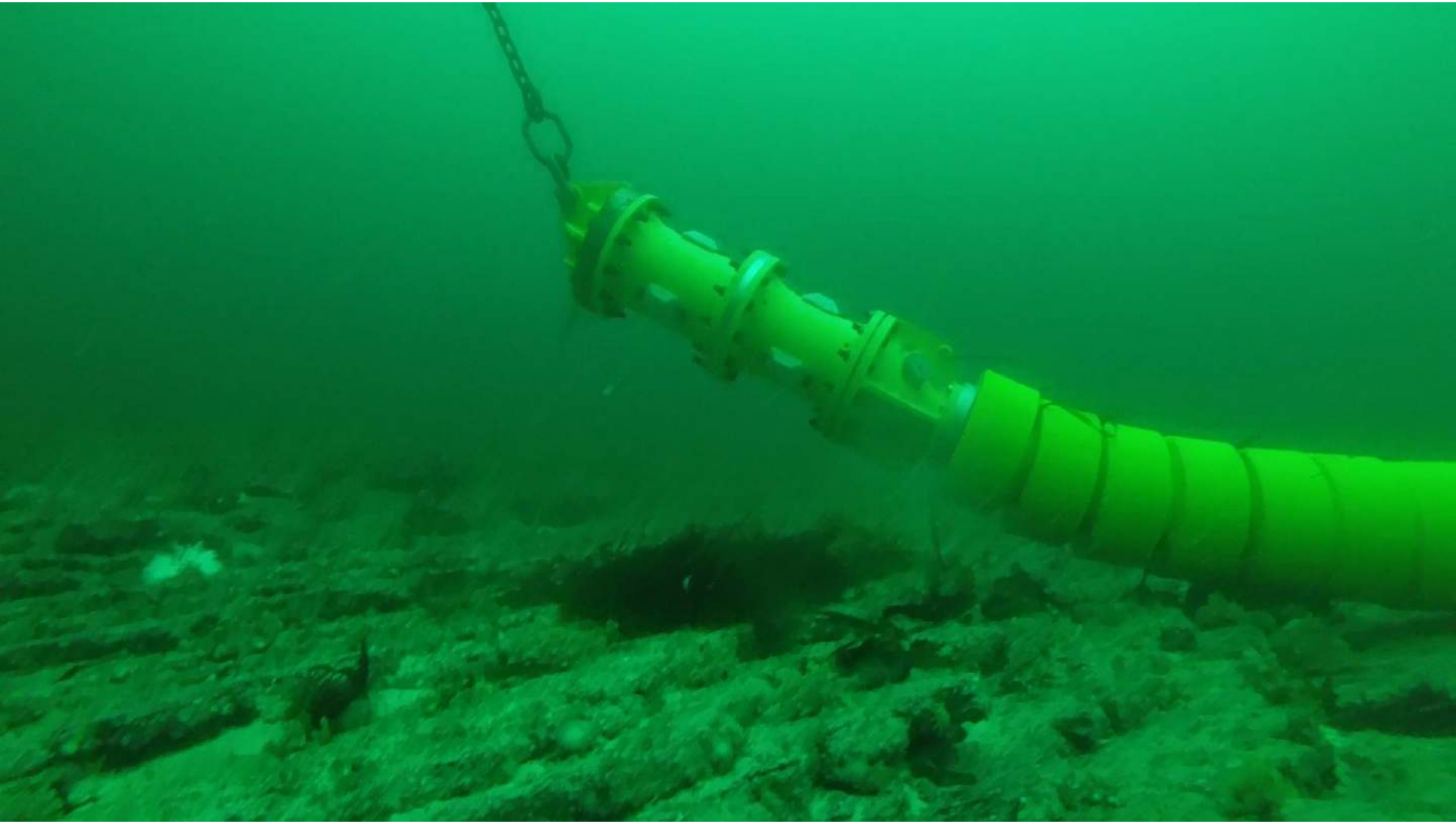
underwater edge

Facts and Figures

- 12.2m length, 2.8m diameter
- Available IT Space: 12 42U racks
- Max Power: 454 KW (38 KW/rack)
- Power Utilization Effectiveness of 1.07
- Payload: 864 Azure servers w/FPGA




700T. SWL



problem space is very rich

data security & integrity

to edge or not to edge?

availability

SLA

resource management

programming model

edge clouds

machine learning
at the edge

economics

serverless framework?

federated edges?

specialized hardware
at the edge

deployments:
drones, automobile, retail, factory floor, homes, enterprise

management
(Kubernetes, …)

5G
cloudification of the telcos

benchmarks

networking

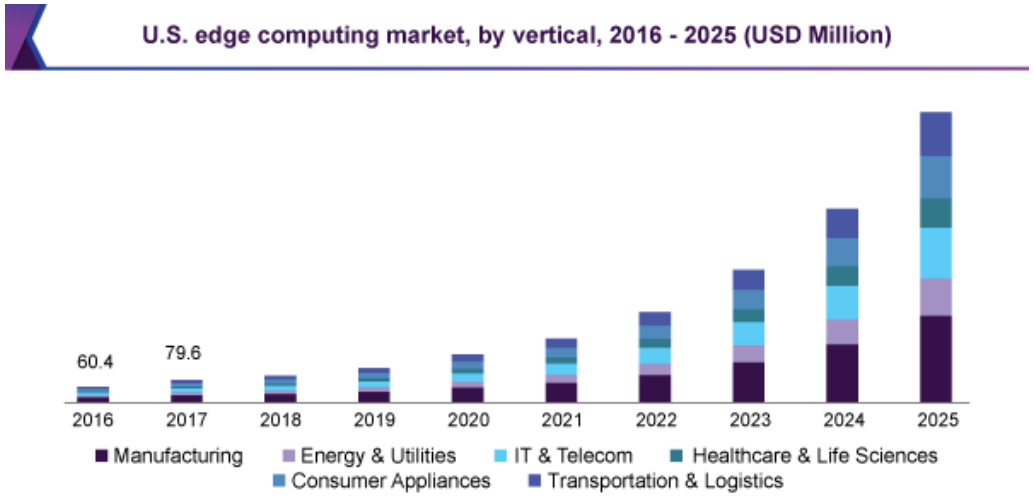
geo-distributed analytics

Microsoft

final thoughts

- edge computing is a paradigm shift, embrace it
  also known as: “micro DCs” & “cloudlets”

- by 2022, video analytics market is expected to become $11.2B and that is going to change lives
  source: *“Video Analytics Market - Global Market Outlook (2016-22)”*, Market Research Consulting Global Inc.



- nation-wide deployments will create a infra-structure where the other aspects of edge computing will shine

Microsoft

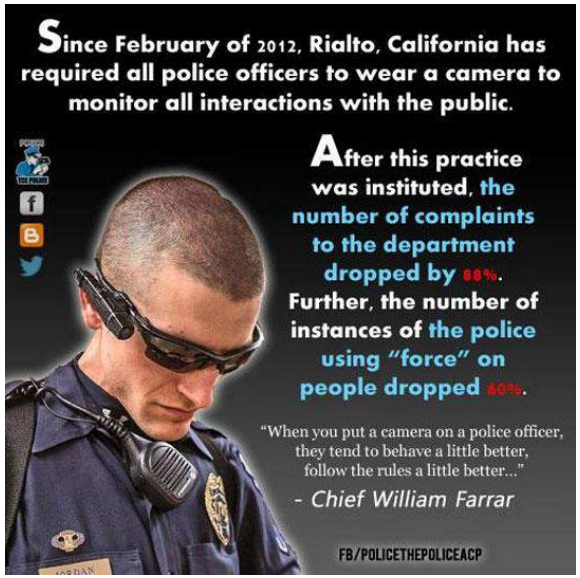
resources

Build with Azure IoT Edge—it's free	https://aka.ms/iot-edge/
get your devices certified on Azure IoT device catalog	https://catalog.azureiotsolutions.com/
nominate your solutions to Azure marketplace	https://aka.ms/iot-edge-marketplace-signup
register for the Vision AI Dev Kit to build your custom vision AI solution	https://www.visionaidevkit.com

Microsoft

thanks!

body worn cameras on the rise



BBC NEWS LONDON updated at 11:11 ET

Metropolitan Police officers start wearing body cameras

The New York Times

New York Police Officers to Start Using Body Cameras in a Pilot Program

By J. DAVID GOODMAN SEPT. 4, 2014

The Washington Post

D.C. police will wear body cameras as part of pilot program

Microsoft