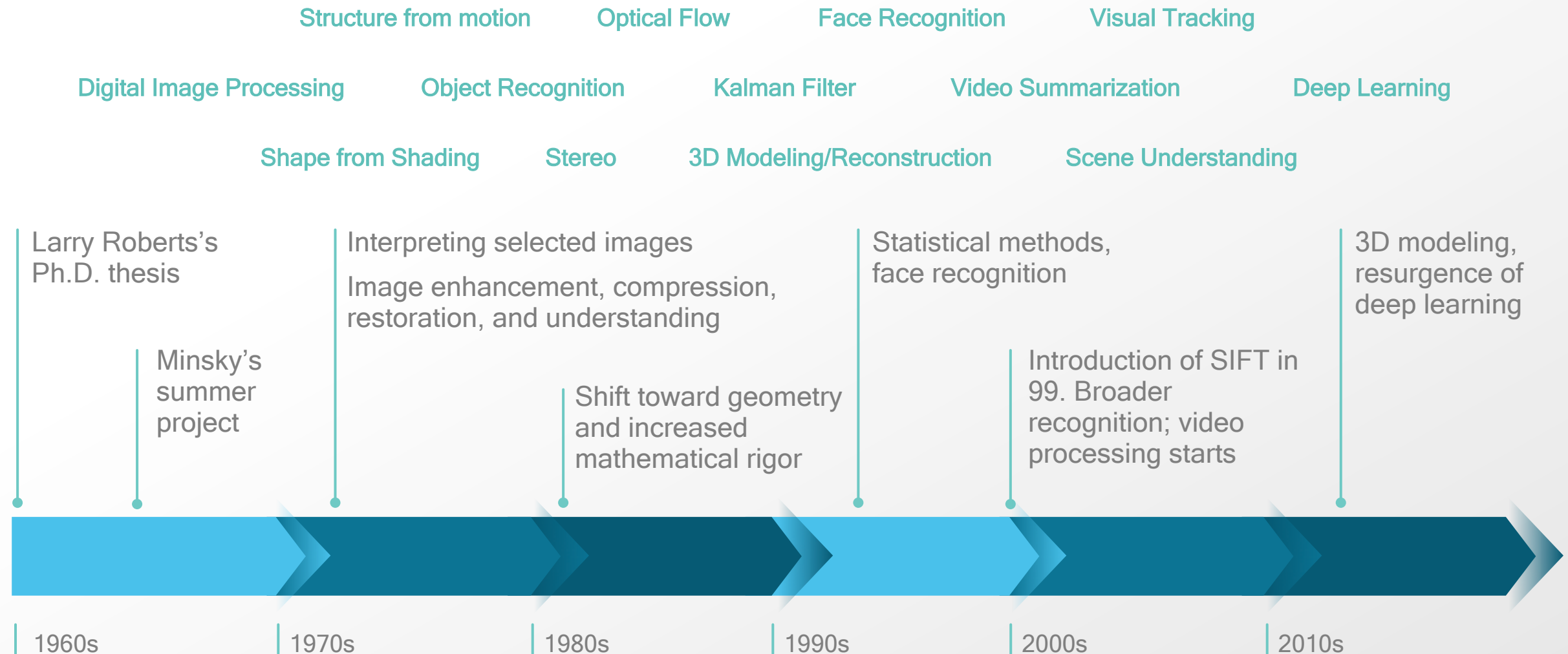# Is vision the new wireless?

Ananth Kandhadai
Senior Director, Engineering, Computer Vision
March 28, 2017

A journey started more than five decades ago
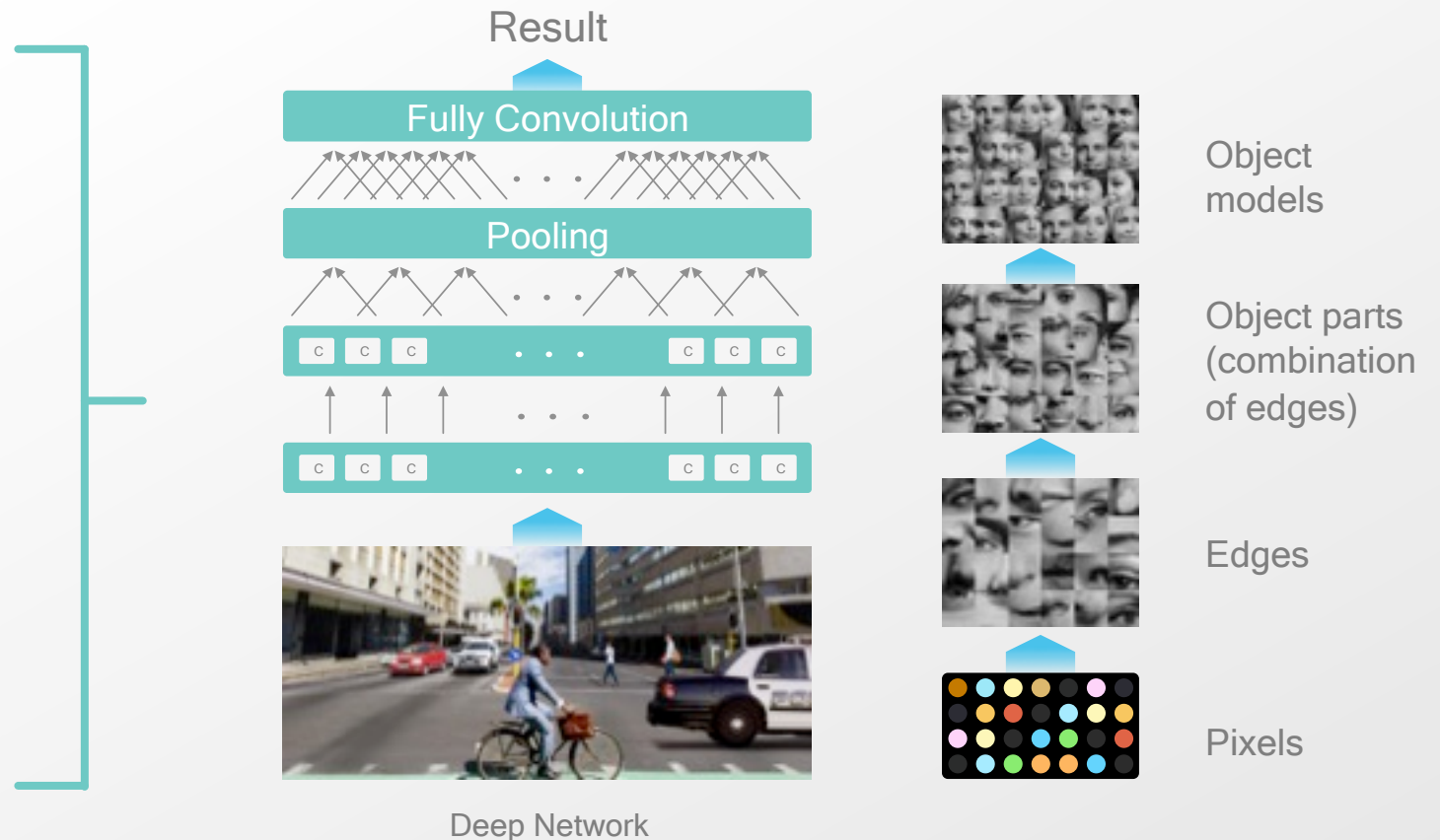
Structure from motion   Optical Flow   Face Recognition   Visual Tracking

Digital Image Processing   Object Recognition   Kalman Filter   Video Summarization   Deep Learning

Shape from Shading   Stereo   3D Modeling/Reconstruction   Scene Understanding

Larry Roberts's Ph.D. thesis

Interpreting selected images

Image enhancement, compression, restoration, and understanding

Statistical methods, face recognition

3D modeling, resurgence of deep learning

Minsky's summer project

Shift toward geometry and increased mathematical rigor

Introduction of SIFT in 99. Broader recognition; video processing starts

1960s   1970s   1980s   1990s   2000s   2010s

Deep Learning techniques have dramatically improved the performance of computer visions tasks such as Object Recognition

Evolving neural network models

- Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for different use cases

- More complex models

Expanding use cases

- ADAS, drones, robotics

- Medical imaging and Genomics

- Security and authentication

Result

Fully Convolution

Pooling

Deep Network

Object models

Object parts (combination of edges)

Edges

Pixels

Intelligence in the mobile and embedded devices is key for ubiquitous use of Vision

Process data closest to the source, complement cloud
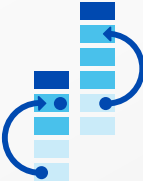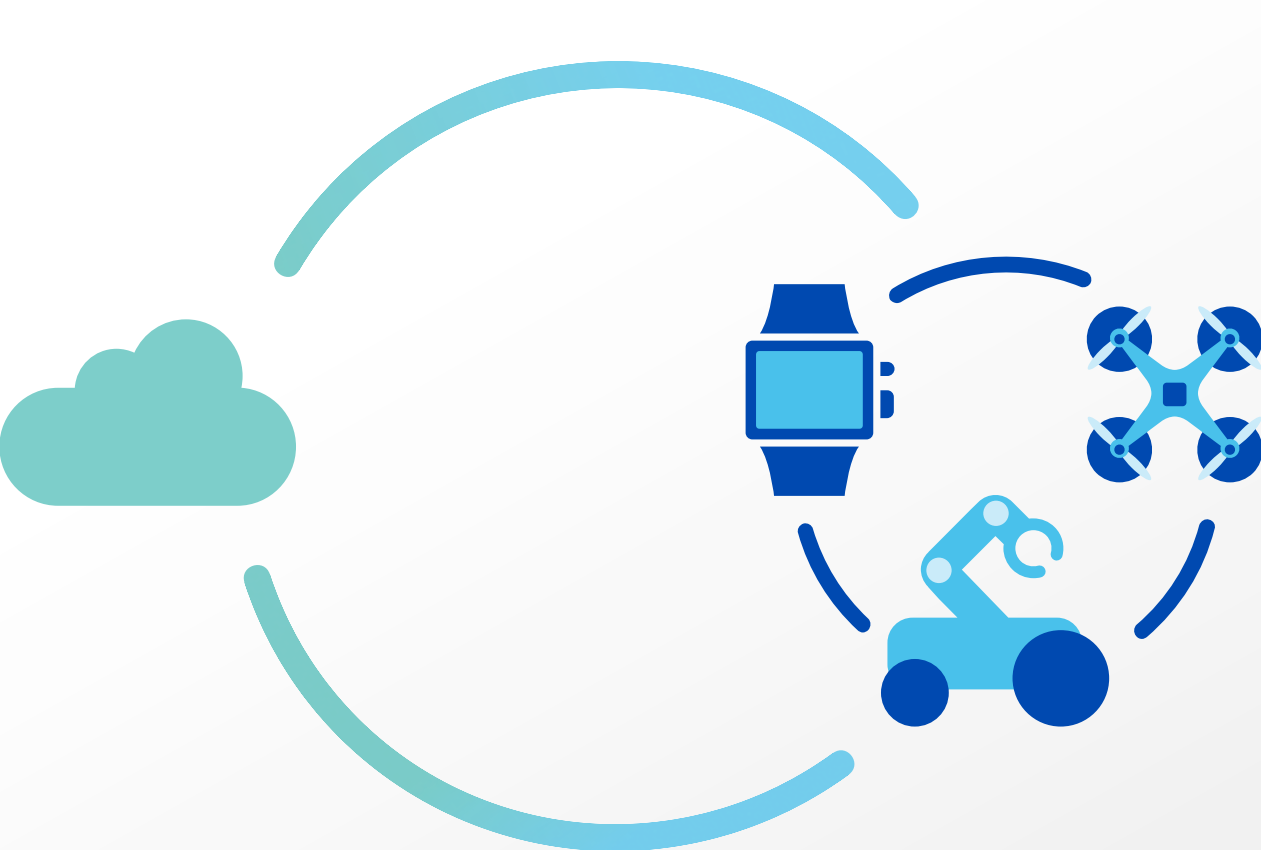
Efficient use of network bandwidth
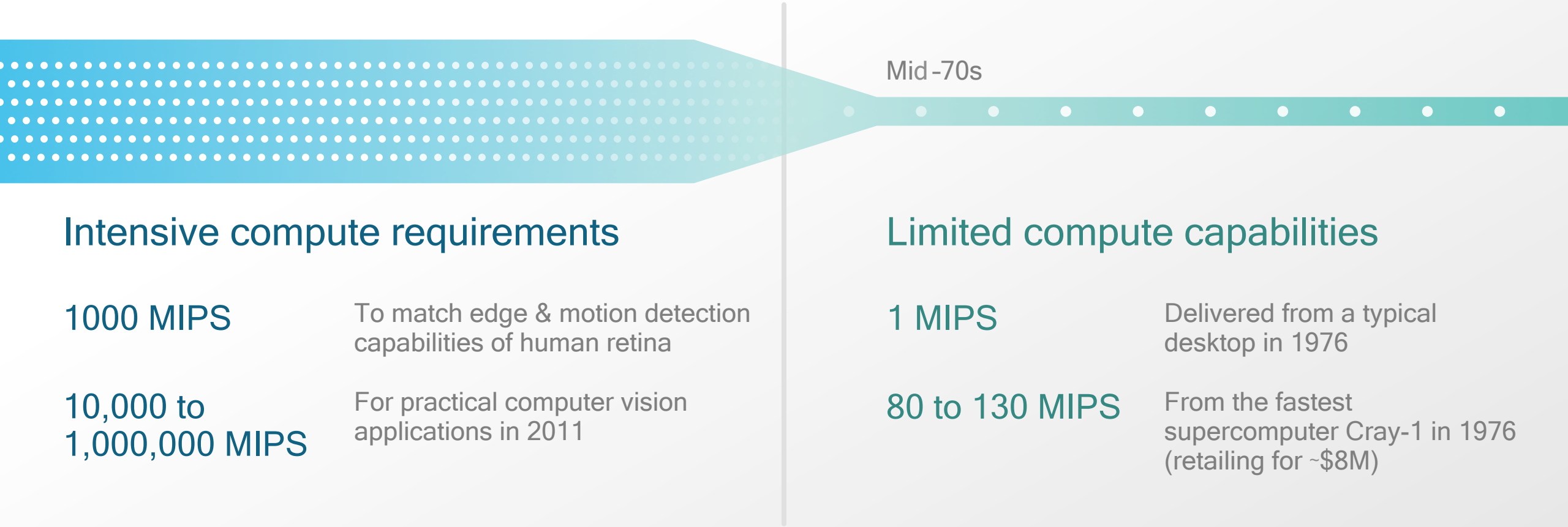
Security and user privacy

Reliability

Low Latency

However, limited compute horsepower and storage slowed the progress

40%-70% of brain capacity used to process visual signals

Mid-70s

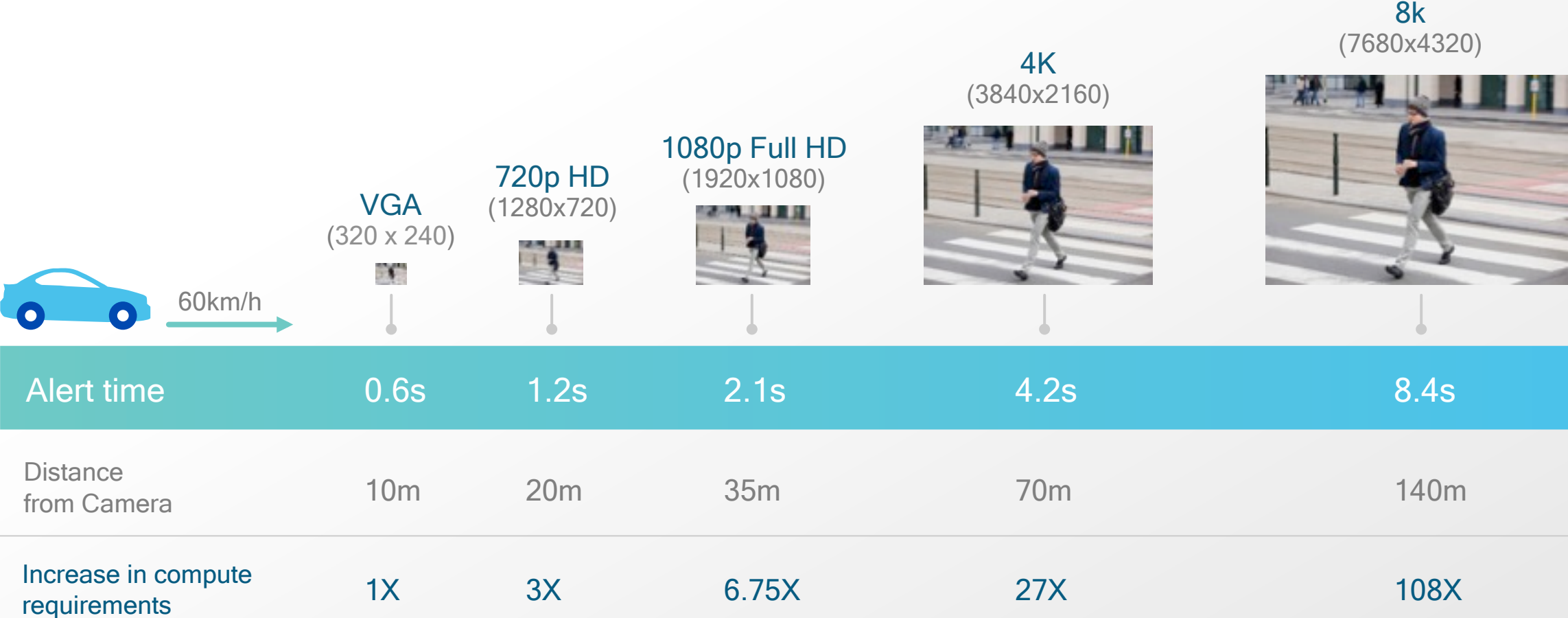### Intensive compute requirements

1000 MIPS	To match edge & motion detection capabilities of human retina
10,000 to 1,000,000 MIPS	For practical computer vision applications in 2011

Limited compute capabilities

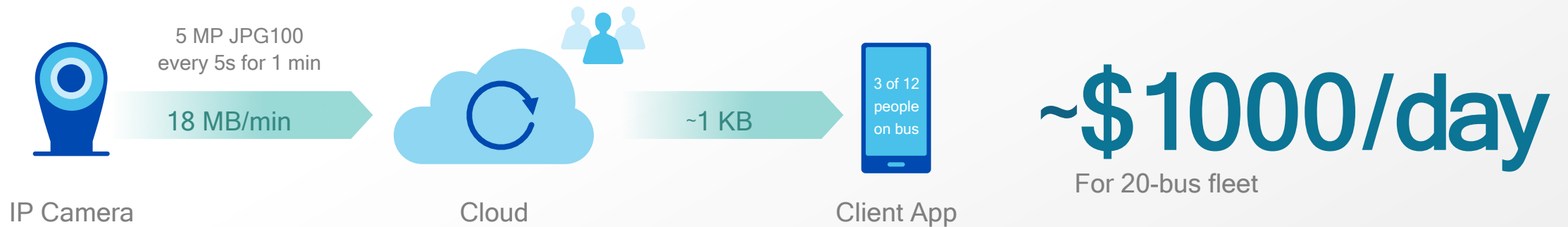|	
1 MIPS	Delivered from a typical desktop in 1976
80 to 130 MIPS	From the fastest supercomputer Cray-1 in 1976 (retailing for ~$8M)

Bringing computer vision to the embedded real world has been challenging

Pedestrian detection: Computer requirements increase with the distance

8k
(7680x4320)

4K
(3840x2160)

1080p Full HD
(1920x1080)

720p HD
(1280x720)

VGA
(320 x 240)

60km/h

	VGA	720p HD	1080p Full HD	4K	8k
Alert time	0.6s	1.2s	2.1s	4.2s	8.4s
Distance from Camera	10m	20m	35m	70m	140m
Increase in compute requirements	1X	3X	6.75X	27X	108X

Example: Intelligence at the edge for IP Camera

Occupancy detection problem

5 MP JPG100
every 5s for 1 min

18 MB/min

IP Camera

Cloud

~1 KB

3 of 12
people
on bus

Client App

~$1000/day

For 20-bus fleet

~1 KB

Intelligent IP Camera

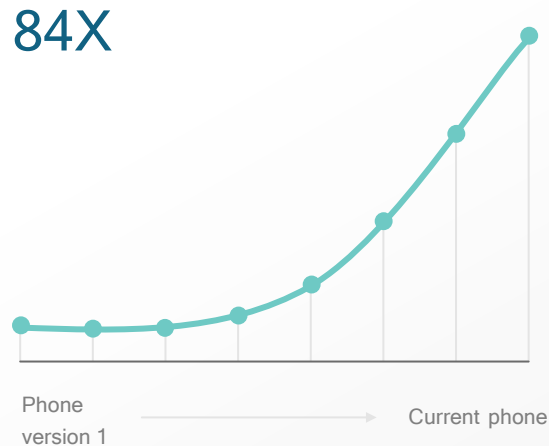Cloud

~1 KB

3 of 12
people
on bus

Client App

~$1/day

For 20-bus fleet

Recent mobile device improvements made it possible

Enhancements in computing, memory, camera sensors and optics

Compute

84X

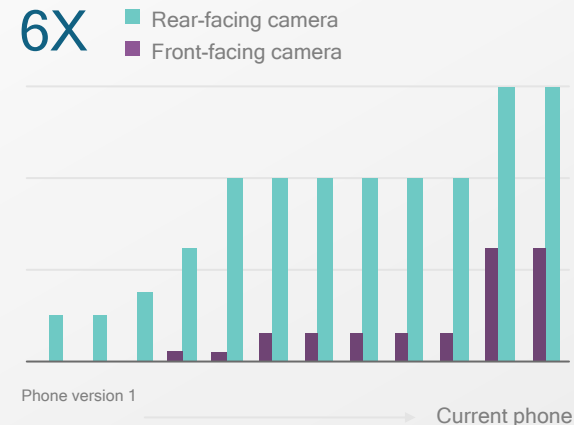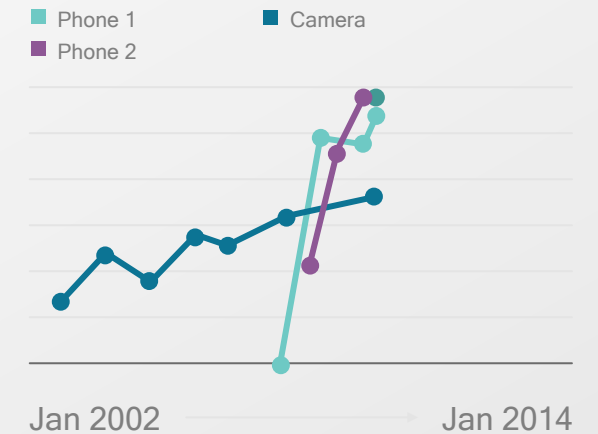Phone version 1 → Current phone

Memory

Smartphone RAM increase (Mb)

16X

2009 → 2016
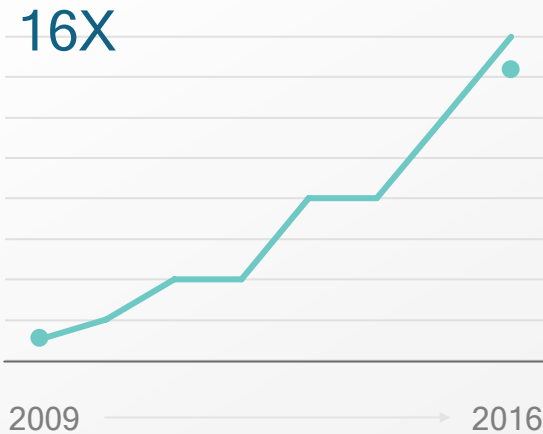
Camera Sensors

Premium camera resolution (stills)

6X

■ Rear-facing camera
■ Front-facing camera

Phone version 1 → Current phone

Optics/overall

Rate of improvements over time in DXOMark

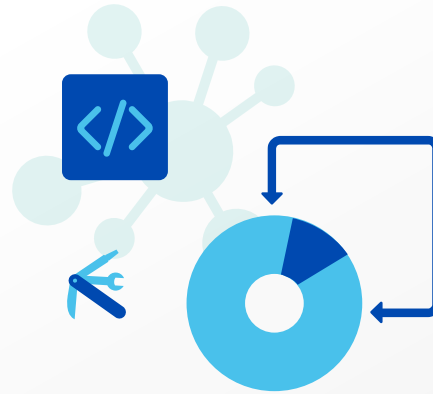■ Phone 1   ■ Camera
■ Phone 2

Jan 2002 → Jan 2014

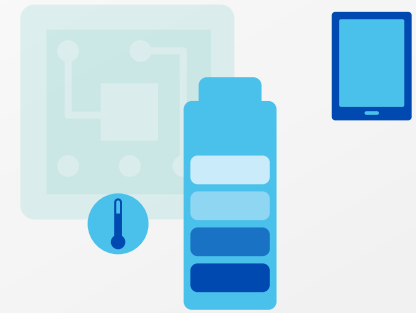# Paving the road of ubiquitous deployment of computer vision

Make technology work in real world

- View point and illumination variations
- Perspective projection, occlusion, deformation and clutter
- Longer distances

Standardization and programmability

- Evolving algorithms and use cases
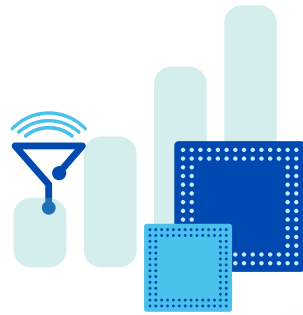- Some emerging standards, software tools and APIs

Bring it to billions of devices

- Mobile and embedded devices have power and thermal constraints

Qualcomm leading the wireless revolution

Investing in wireless for many years—building upon our leadership foundation

Wireless technologies and chipset leadership

Pioneering technologies to meet extreme requirements

End-to-end system approach with advanced prototypes

Applications Processing, Various forms of Connectivity Driving standardization to commercialization

Leading global network experience and scale

Providing the experience and scale that meets world wide demands

Qualcomm® Snapdragon™ Neural Processing Engine SDK

Software accelerated runtime for the execution of deep and recurrent neural networks

Smartphone

IP Camera

Automotive

Drones

IoT

Snapdragon Neural Processing Engine

Released for Snapdragon last year

Efficient execution
- Taking advantage of Snapdragon's heterogeneous computing capabilities

Common model framework support
- Convolutional or recurrent neural networks
- Support of Caffe and Cuda-convnet

Debugging and optimization tools
- Debug and analyze network performance
- Ease of integration into customer applications

Computer vision is empowering broad set of applications

IP Camera

Robotics

VR/AR

Drones

Mobile imaging

Automotive
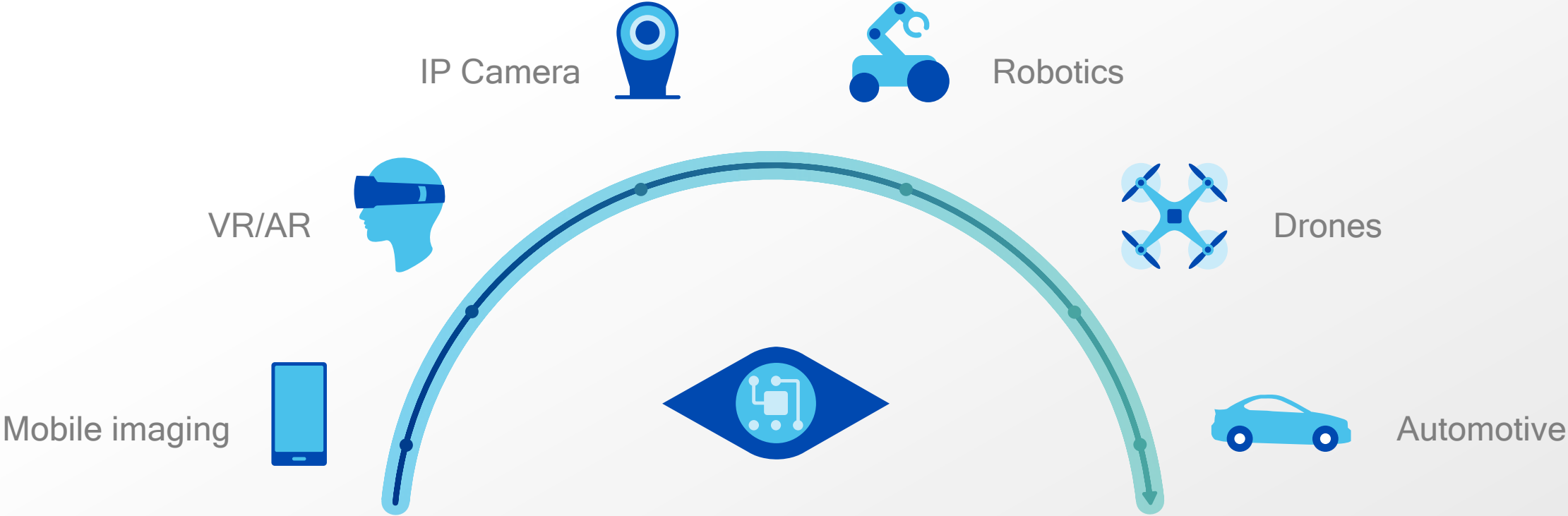
Already creating breakthrough mobile experiences in billions of devices

- Biometric security

- Face beautification

- Clever capture

- Text activation

- Face/smile detection

8.7B*

*cumulative smartphone shipments estimated from 2016 -2020

Qualcomm

- Low-light photography

- Optical-like zoom

- Touch-to-track

- Panoramic stitching

- Eye/head tracking

Vision will be one of the key enabling technologies for both AR and VR

Hand gesture
Using computer vision for command and control

Head and eye tracking
By using Visual-Inertial Odometry (VIO) for accurate 6-DOF pose

360 spherical view
Undistort, calibrate, stitch together, and map the discrete images to a equirectangular or cube map format

Foveated rendering
To significantly reduce pixel processing with the help of eye tracking

Vision and video analytics are redefining the connected camera experiences

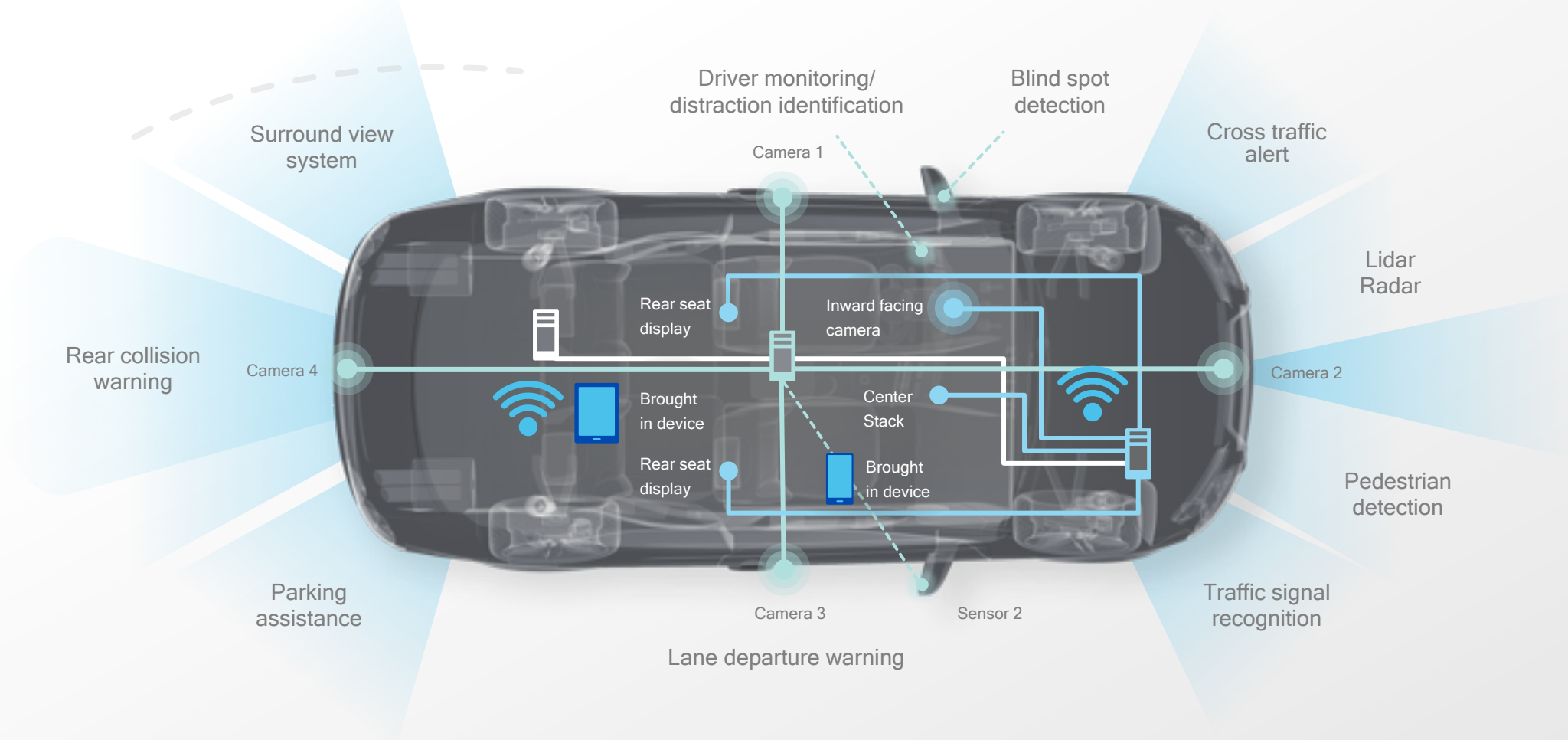**Home Surveillance**
Keeping our homes safer

Action Camera
Capturing our important moments

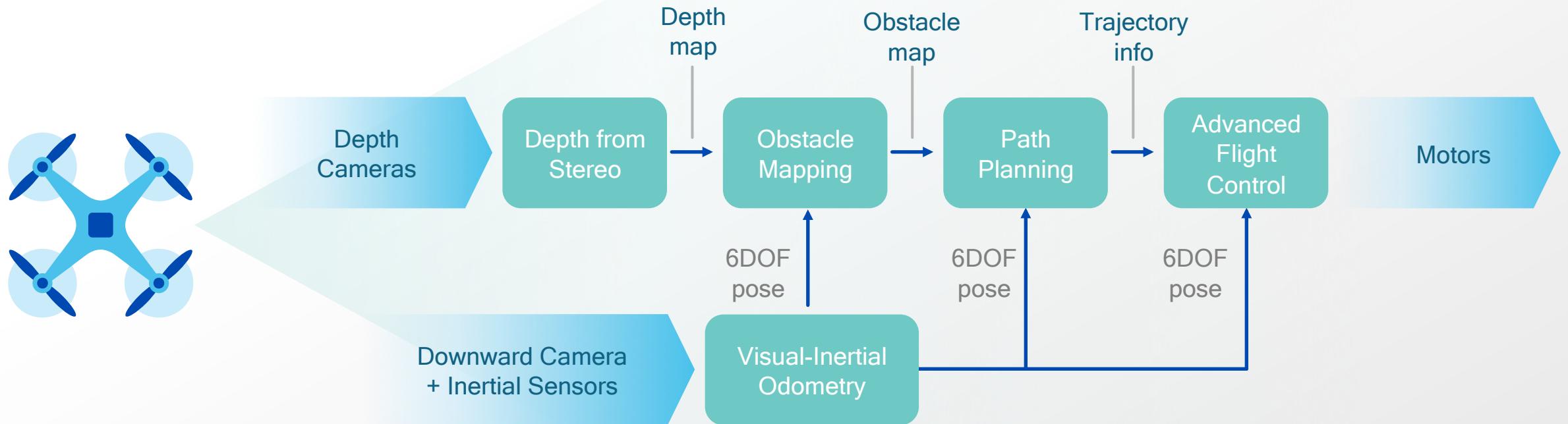**Professional Surveillance**
Keeping our communities safer

Vision is enabling ADAS today and autonomous driving in the future

ADAS demand to hit $19.9B by 2020 with CAAGR of 19.2% over 2015 to 2020
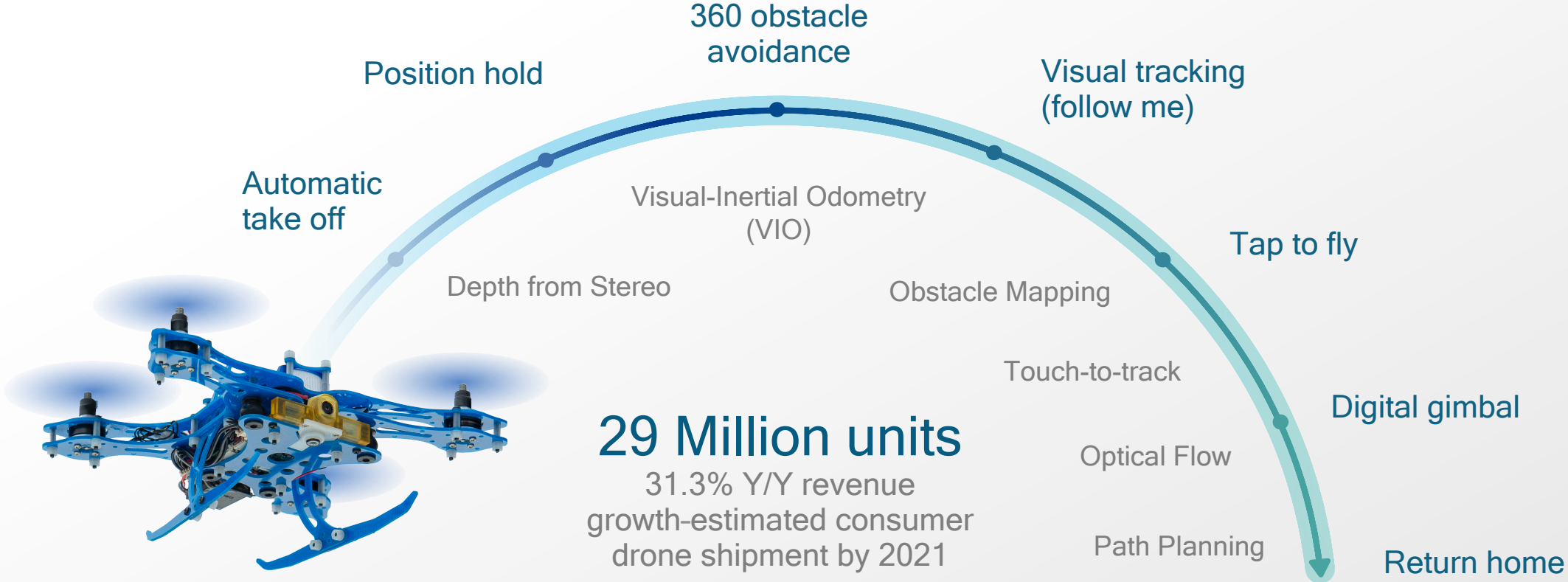

Surround view system

Driver monitoring/distraction identification

Blind spot detection

Cross traffic alert

Camera 1

Lidar Radar

Rear seat display

Inward facing camera

Rear collision warning

Camera 4

Camera 2

Brought in device

Center Stack

Rear seat display

Brought in device

Pedestrian detection

Parking assistance

Camera 3

Sensor 2

Traffic signal recognition

Lane departure warning

Autonomous visual navigation for drones and robotics

Depth Cameras → Depth from Stereo → (Depth map) → Obstacle Mapping → (Obstacle map) → Path Planning → (Trajectory info) → Advanced Flight Control → Motors

Downward Camera + Inertial Sensors → Visual-Inertial Odometry → 6DOF pose → Obstacle Mapping; 6DOF pose → Path Planning; 6DOF pose → Advanced Flight Control

Enabling advanced use cases for millions of consumer drones

By using computer vision and machine learning

360 obstacle avoidance

Position hold

Visual tracking (follow me)

Automatic take off

Visual-Inertial Odometry (VIO)

Tap to fly

Depth from Stereo

Obstacle Mapping

Touch-to-track

Digital gimbal

29 Million units
31.3% Y/Y revenue growth–estimated consumer drone shipment by 2021

Optical Flow

Path Planning

Return home

Ubiquitous deployment of visual intelligence enables contextual awareness

Bee-sized flying cameras

Intelligent cameras

Adaptive self-driving cars

Aerial, 360 Virtual reality

Closing Thoughts

- Computer Vision is a fundamental technology that has the potential to change the way we live and use machines

- After many decades of research it is now finally beginning to show great results

- Embedded vision is key to ubiquitous adoption of vision in our daily life

- Exciting times ahead us as we deliver mobile platforms and technologies for mass scale deployment of computer vision–in harmony with the cloud

Thank you

Follow us on: **f** 🐦 **in** **t**

For more information, visit us at:
www.qualcomm.com & www.qualcomm.com/blog