

Time-Critical Active Learning for Edge-based Video Analytics

Mahadev Satyanarayanan
School of Computer Science
Carnegie Mellon University

Credits

**Shilpa George, Haithem Turki, Ziqiang Feng,
Padmanabhan Pillai (Intel), Deva Ramanan, Mahadev Satyanarayanan**

Video Sensing: Key Driver of Edge Computing

Unique attributes

- **non-invasive** (as opposed to embedded sensors)
- **very high resolution**
- **large coverage area**
- **flexibility after installation** (new video analytic algorithms)
- **direct comprehensibility by humans** (skip the ML-based inference step)



Missing Child



Icy Sidewalk



Long Checkout Line



Wasted Advertising



Spilled Liquid

24x7x365 Flood of Inbound Video Data

No way to ship all captured data to cloud

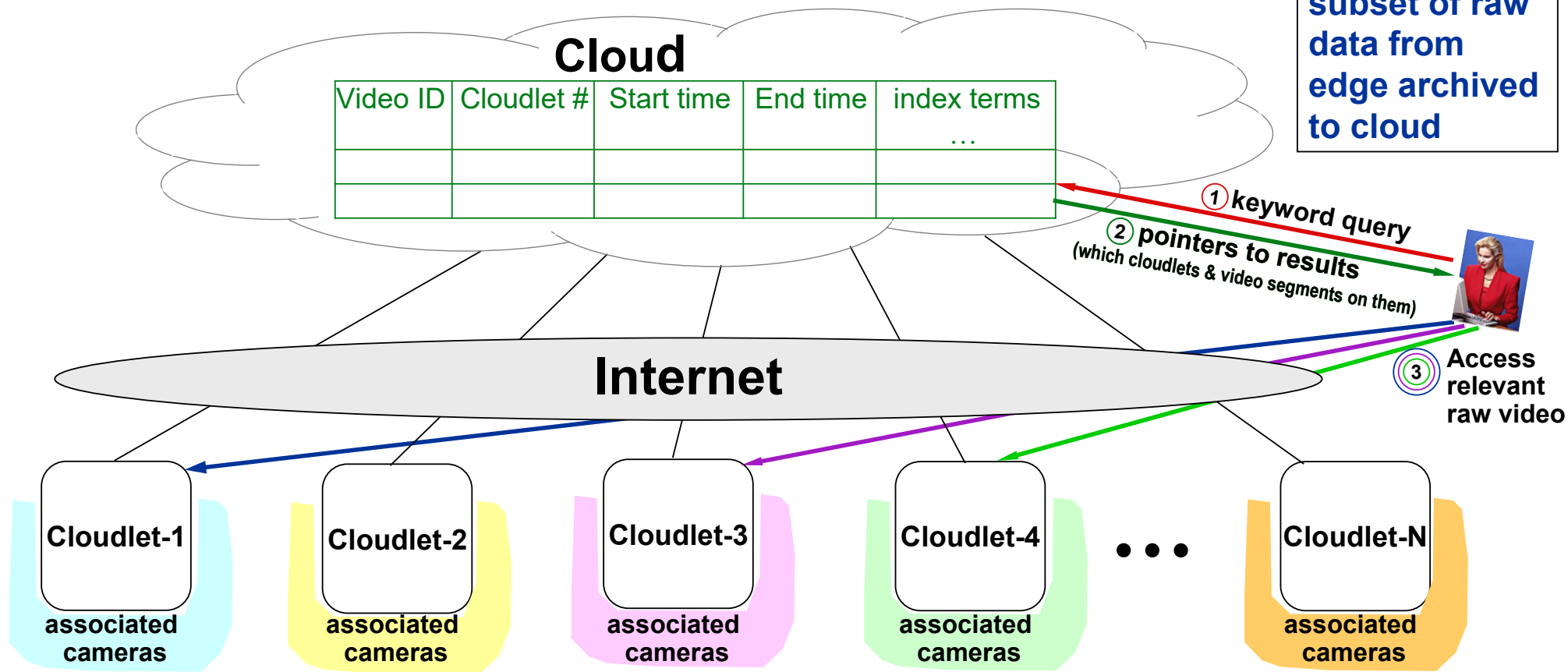
Has to be processed at edge (on cloudlet)

(eventually discarded if unused after long time, days/weeks/months)

What can be shipped to cloud

- **extracted information, with raw data annotation**
- **tiny fraction of raw data for archiving/training**

Searching with Pre-built Index



Finite Data Retention on Cloudlets

Indexed Search \equiv Context-Free Search

Classical indexing can only go so far

- enormous space of possibilities
- knowledge base at time of indexing and search may be different
- *person performing search may possess unique knowledge*
- index was created without this knowledge

High-value searches are often context-sensitive

- existing indexes only modestly narrow search space
- the “last mile” of the search involves a lot of raw data
- slow and painful
- can we help?

Example: Lost Dog

Missing for 24 hours

Where do we look? Is the dog on some video? City allows search

- dog could be anywhere within an area of 12 square miles in 24 hours
half the size of Manhattan, @ 2 mph
- over 100,000 surveillance cameras outside residences and businesses
London estimated to have 500,000 surveillance cameras today
- 1 frame indexed every 10 seconds per camera → 800 million frames in 24 hours

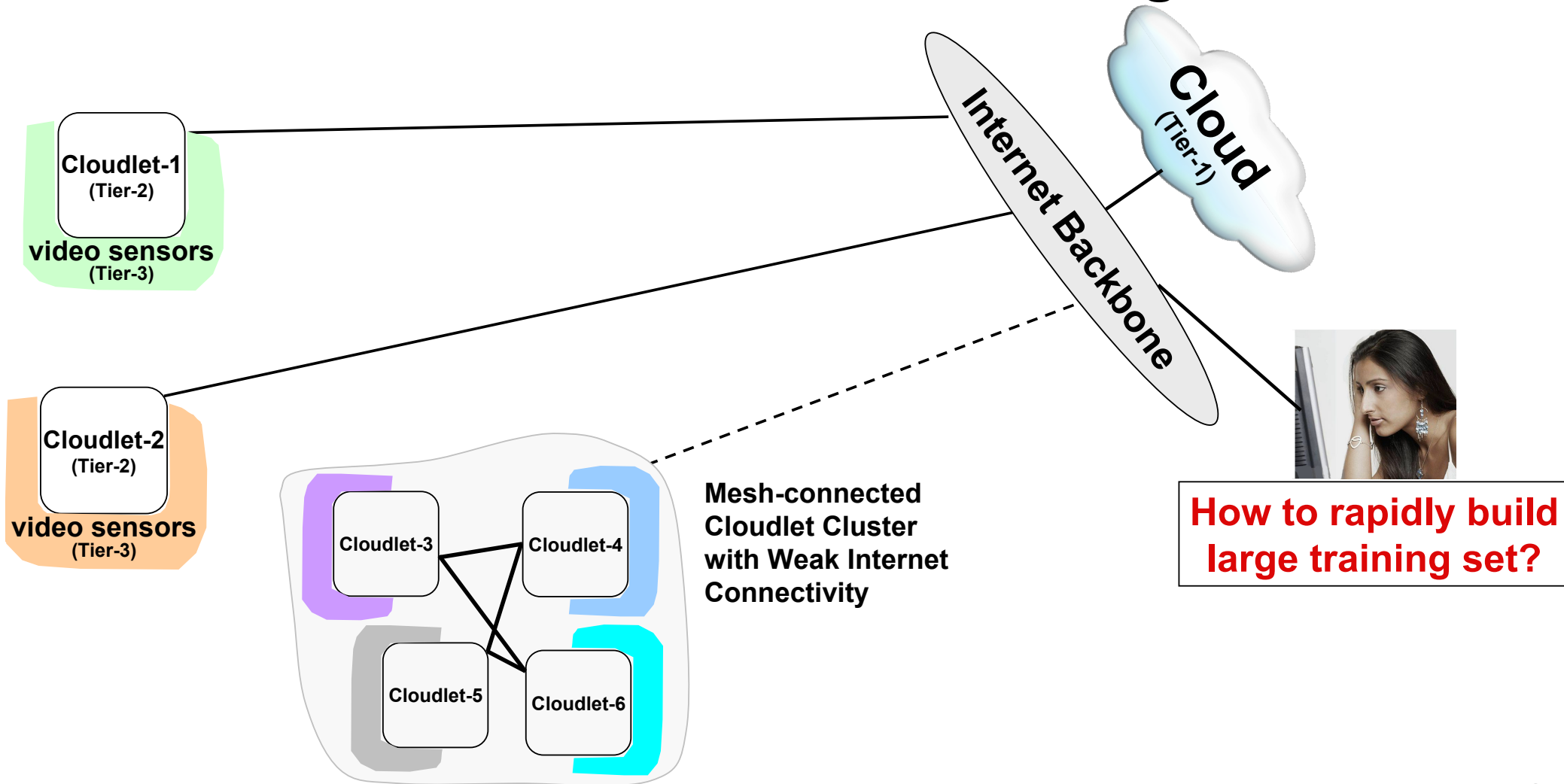
Index term “dog” shrinks space considerably

- in a dog-friendly city, 0.1% hit rate reasonable → 800,000 frames with dogs
- is any of them my dog? still a daunting task

Need for tools that allow context-sensitive search of video data



Create Classifier for New Target



Key Challenges

Need to find many positive examples for training set

- 10^2 to 10^3 examples for transfer learning
- 10^1 to 10^2 examples even for SVM on DNN features

What makes this difficult?

1. **Low base rate:** very few hits in a huge mass of data (highly skewed distribution)
2. **Context sensitivity:** hits only likely on new edge data (can't find it in the cloud)
3. **Low bandwidth:** can't fetch all the data from cloudlets
4. **Expertise & Classified data:** can't crowd-source (only authorized experts can label data)

Our Approach

Principle 1: Use Edge Computing to do “early discard” (bandwidth savings)

Principle 2: Use Edge Computing to do learning (negative + positive examples are at the edge)

Principle 3: Seamlessly integrate learning and labeling (frequently deploy new, improved classifier)

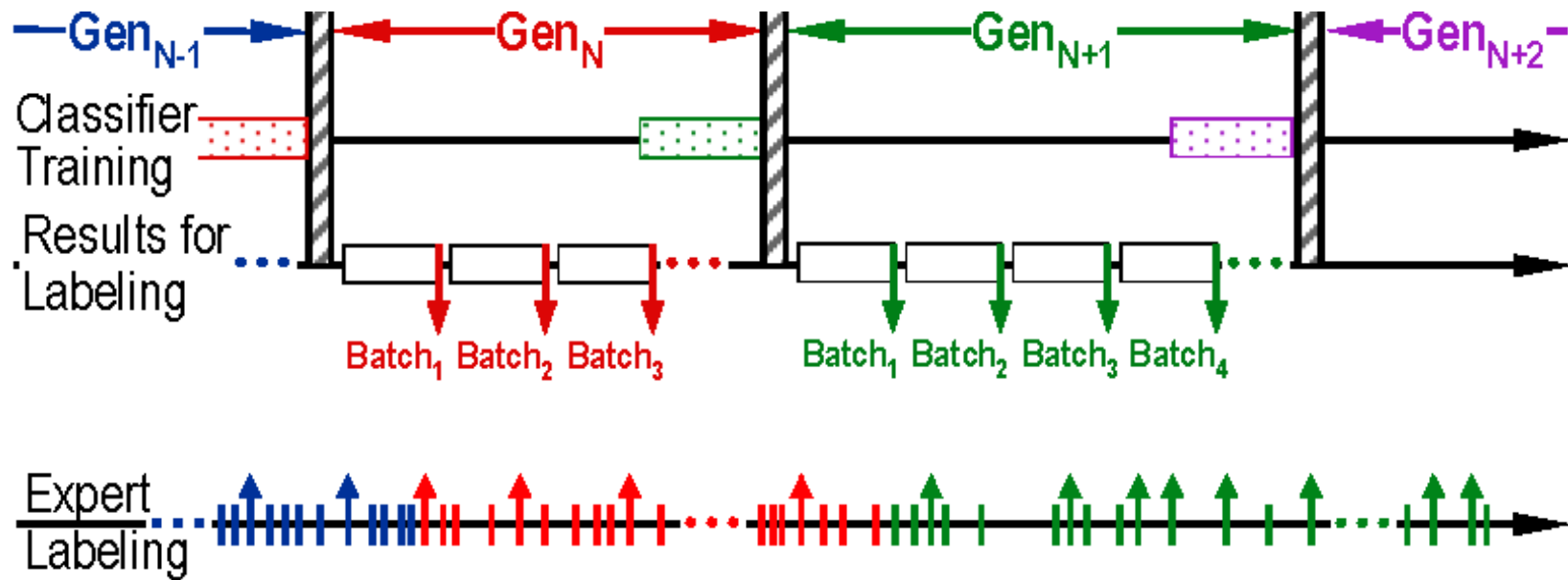
Principle 4: The Human Expert is the most precious resource

- avoid wasting time on frivolous examples
- minimize stalls, waiting for work

Prototype Implementation: Delphi

Transparent, Recursive, Scalable Learning

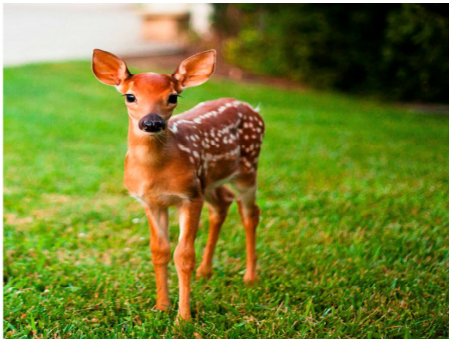
Occurs in Parallel on Each Cloudlet



Case Study

100 Million Image YFCCOM Dataset

Training Example



Discovered Examples



Target	Images Processed	Images Shown	Posi- tives Found	Elapsed Time (min)	Positives Found / Elapsed Time
Deer	534,869 (144,704)	1,041 (261)	101 (3)	12.55 (3.13)	8.49 (1.69)
Taj Mahal	1,866,285 (165,919)	3,780 (350)	103 (2)	40.50 (3.26)	2.55 (0.22)
Fire Hydrant	2,269,136 (174,916)	4,524 (381)	103 (2)	49.03 (3.34)	2.10 (0.14)

In Closing ...

Edge computing simplifies

- capture of data
- preprocessing data
- retention for finite period

How do we extract value from this data?

Delphi: just-in-time human-guided edge-based learning