Scaling Video Analytics to Large Camera Deployments

Samvit Jain, Ganesh Ananthanarayanan,
Junchen Jiang, Yuanchao Shu, Joseph Gonzalez

Microsoft Research
University of California, Berkeley
University of Chicago

Trends

- Falling camera costs
  - **2009** – 780p IP camera (Axis) – **$1,500**
  - **2019** – 1080p IP camera (Wyze) – **$20**

Axis 720p IP camera (2009)

Wyze 1080p IP camera (2019)

Trends

- Advances in computer vision
  - Progress on benchmarks
    - **Image** – classification, object detection
    - **Video** – action recognition, object tracking
  - New capabilities
    - Identity verification
    - Autonomous navigation
    - Visual monitoring

KLM's biometric boarding pass

Trends

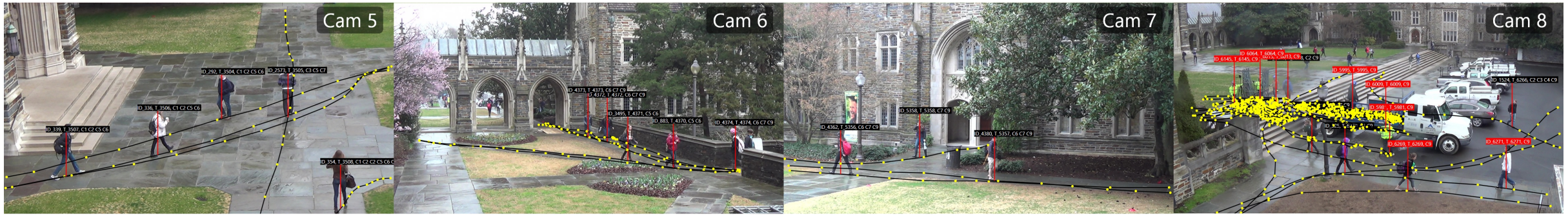
- Rise of large, centralized video analytics operations
  - **London** – 12,000 cameras on rapid transit system
  - **Chicago** – 30,000 cameras across city
  - **Paris** – 1,500 cameras in public hospitals

Cross-camera analytics

- Problem statement
  - Given: instance of query identity **Q**
  - Return: all later frames in which **Q** appears
- Application space
  - Real-time search
    - Track threat (e.g. AMBER alert)
  - Post-facto search
    - Investigate crime (e.g. terrorist attack)
  - Trajectory analysis
    - Learn customer behavior (e.g. in a supermarket)

Large camera deployments

- Challenges
  - **High inference cost** ($$)
    - Example: Chicago Public Schools' deployment of 7000 security cameras
      - $28 million in GPU hardware (at $4,000 / GPU)
      - $1 million/month in GPU cloud time (at $0.9 / GPU hour)
  - **Low inference accuracy** (R/P, F1)
    - Tasks of interest (anomaly detection, instance retrieval) are difficult



Approach: collaborative analytics

- Leverage **cross-camera correlations** to:
  - Constrain tracking search space → lower cost
  - Decrease frequency of false matches → higher accuracy

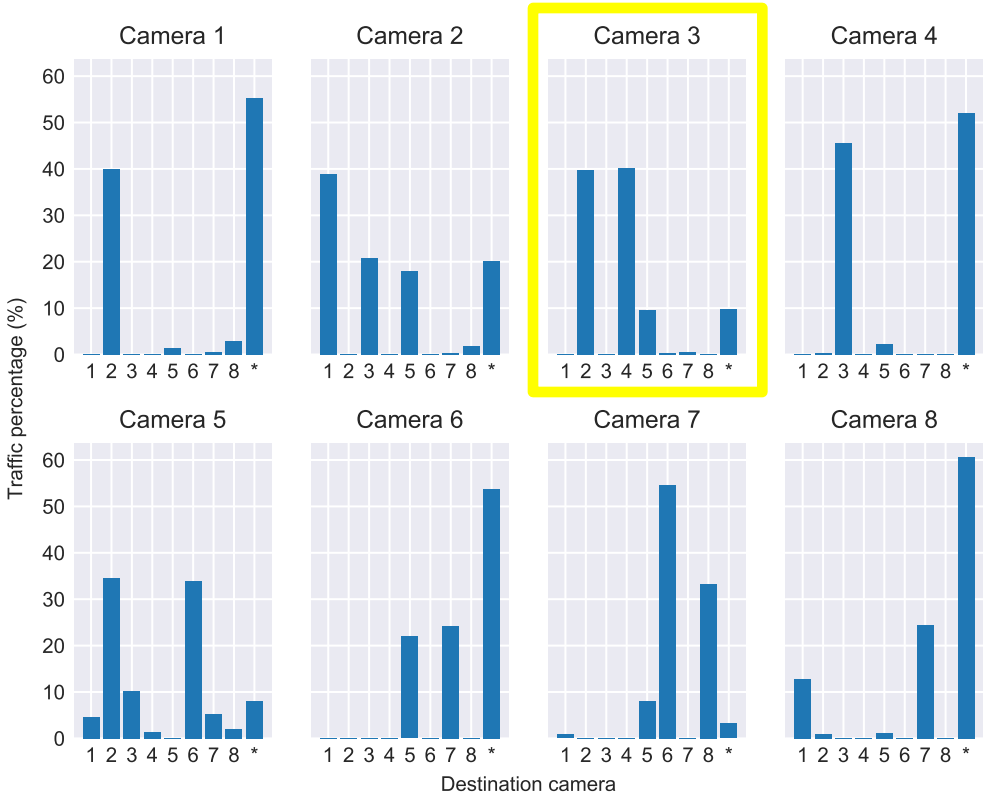
- Types of correlations:
  - **Spatial** – which **cameras** are likely to receive outgoing traffic from camera X
  - **Temporal** – which **frames** are likely to contain outgoing traffic from camera X

Cross-camera correlations

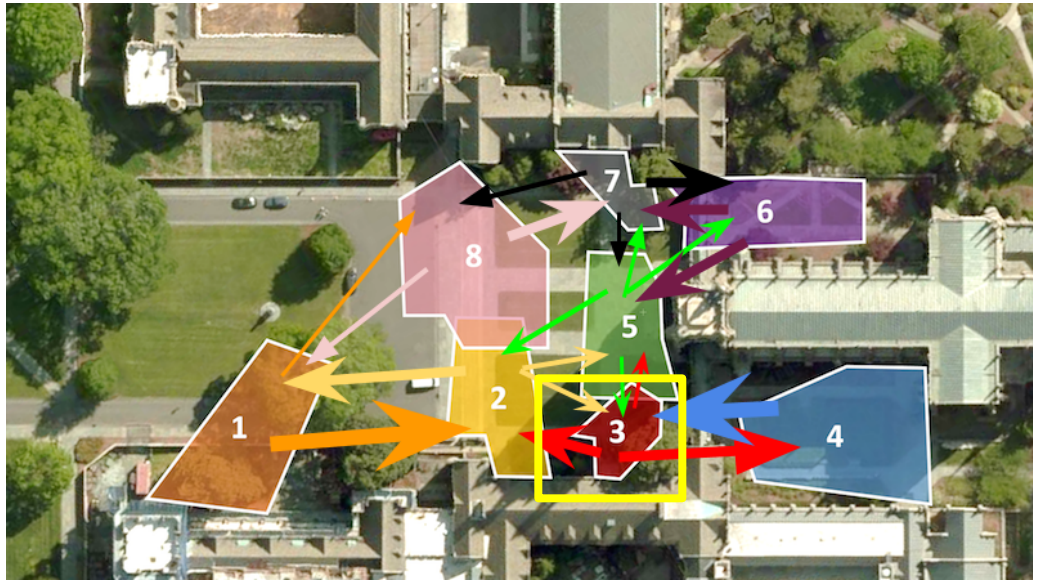
- Dataset
  - 1080p video from 8 cameras
  - 2 million annotated frames
  - 2,700 unique identities

Person detections in DukeMTMC dataset

Cross-camera correlations

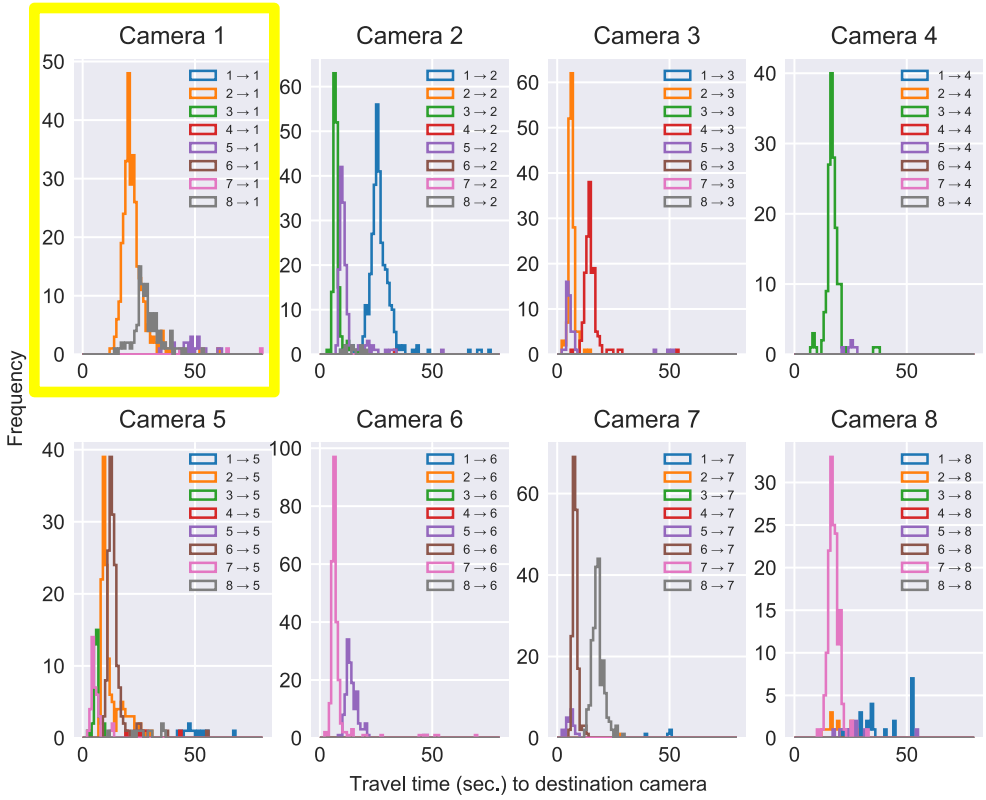


Spatial correlations



Spatial correlations (flow graph)

Cross-camera correlations



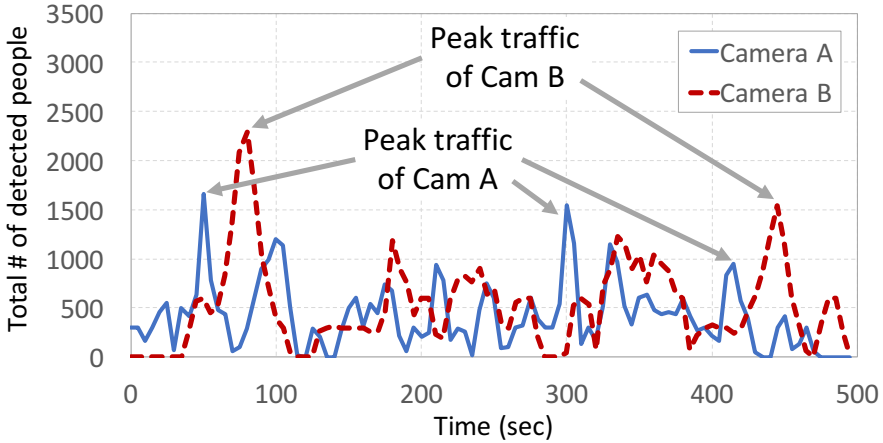
Temporal correlations

Results: spatio-temporal pruning

Filtering level (%)	Detections processed	Savings (vs. baseline)	Recall (%)	Precision (%)
0%	76,510	--	**57.4**	60.6
1%	29,940	2.6x	55.0	81.4
3%	22,490	3.4x	55.1	**81.9**
10%	19,639	**3.9x**	55.1	**81.9**

Opportunities

- Lower inference cost ($$)
  - **Spatio-temporal pruning**
  - Resource pooling
- Higher inference accuracy (R/P, F1)
  - **Spatio-temporal pruning**
  - Cross-camera model refinement



Person detections on Cameras A, B over time

Questions for the future

- **How** will the multi-camera analytics stack be **architected**?
  - Cross-camera correlations
  - Data and model sharing
  - Synchronous vs. asynchronous