Information Processing Techniques for Mobile Sensing Applications *at* Scale

ACM India Summer School 2025: AI for Social Good
Hosted by IIT Gandhinagar

Prasant Misra

W: https://sites.google.com/site/prasantmisra/
W: https://www.linkedin.com/in/prasantmisra/

City Scale Monitoring of On-Street Parking Violations

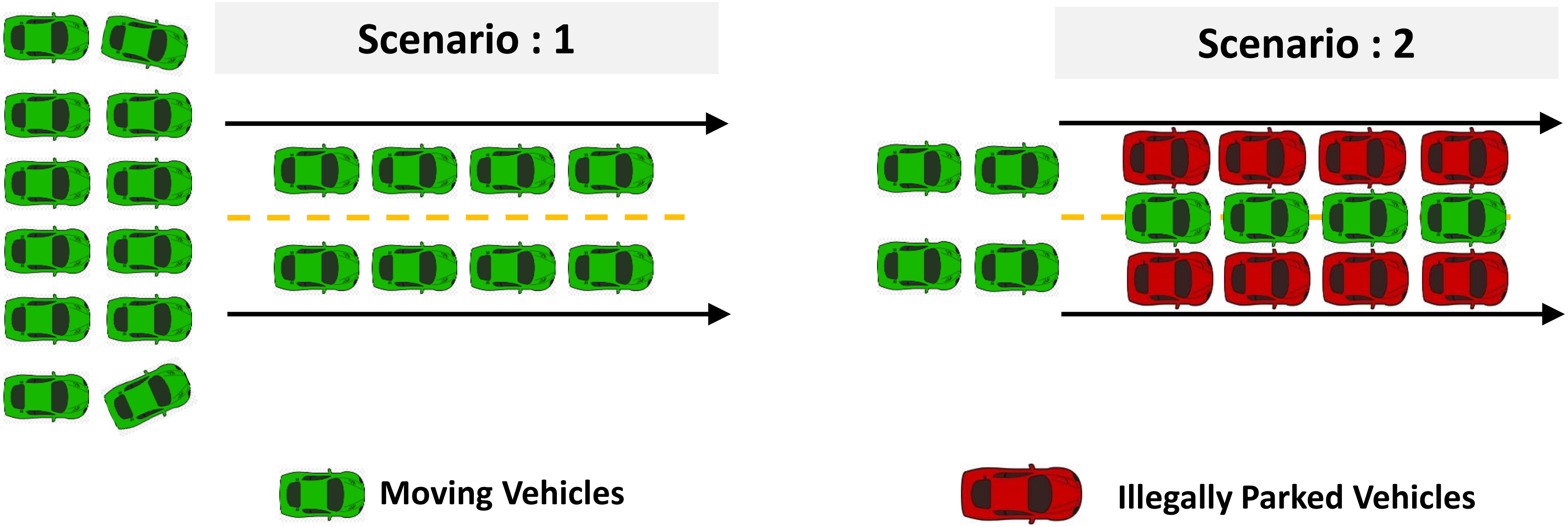
(Published in ACM BuildSys'19)

A. Ranjan, P. Misra, A. Vasan, S. Krishnakumar, and A. Sivasubramaniam, "City Scale Monitoring of On-Street Parking Violations with StreetHAWK", In *Proceedings of the 6th ACM International Conference Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys'19, pages 31-40, NY, USA, 2019. ACM.

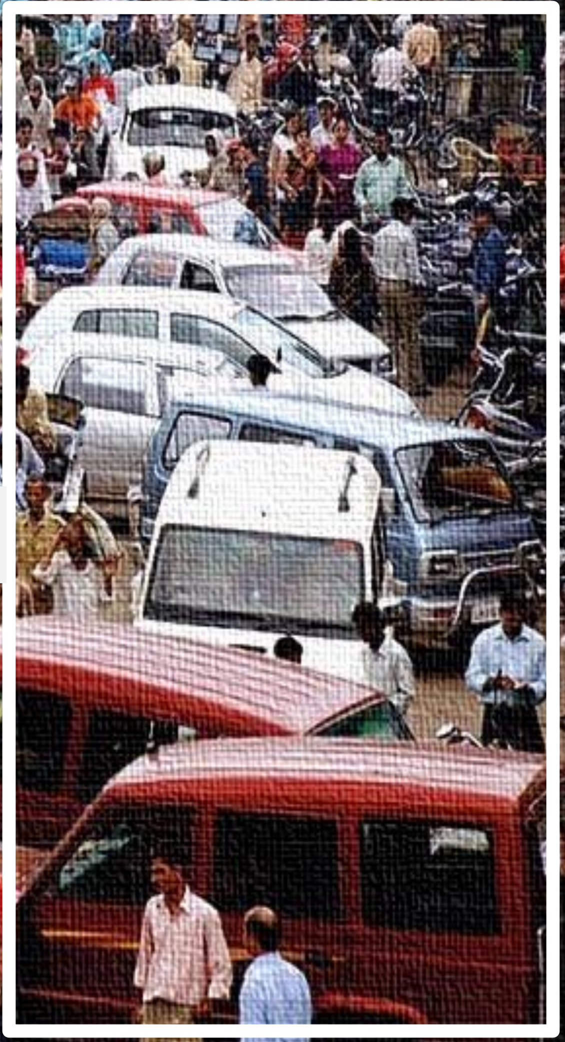
Slow Moving Traffic | Traffic Congestion | Developing Countries

Traffic Congestion and On-street Parking Violations?

Congestion := Demand exceeds Capacity



Illegal parking takes up 40% of road/street space!



Problem Statement

How to **monitor** **illegal** **parking** practices at **city scale** with **limited** human intervention?

State of Current Practice

How is it done now?

- Focus is on identification and deterrence!
- Identification best practices
  - (Semi) manual analysis of video feeds from traffic surveillance cameras
  - Patrolling by enforcement agencies
- Deterrence procedures
  - Issue of violation tickets with monetary fines; towing; wheel clamping

Limitations of current practice?

- Monitoring is (mostly) manual
- Difficult to scale-up and maintain the monitoring system with problem size; and is costly
- Streaming visual data to the cloud has privacy implications

The StreetHAWK Proposal

Piggyback on the ubiquitous smartphone infrastructure that is readily available on radio taxis moving in the city!

The StreetHAWK System

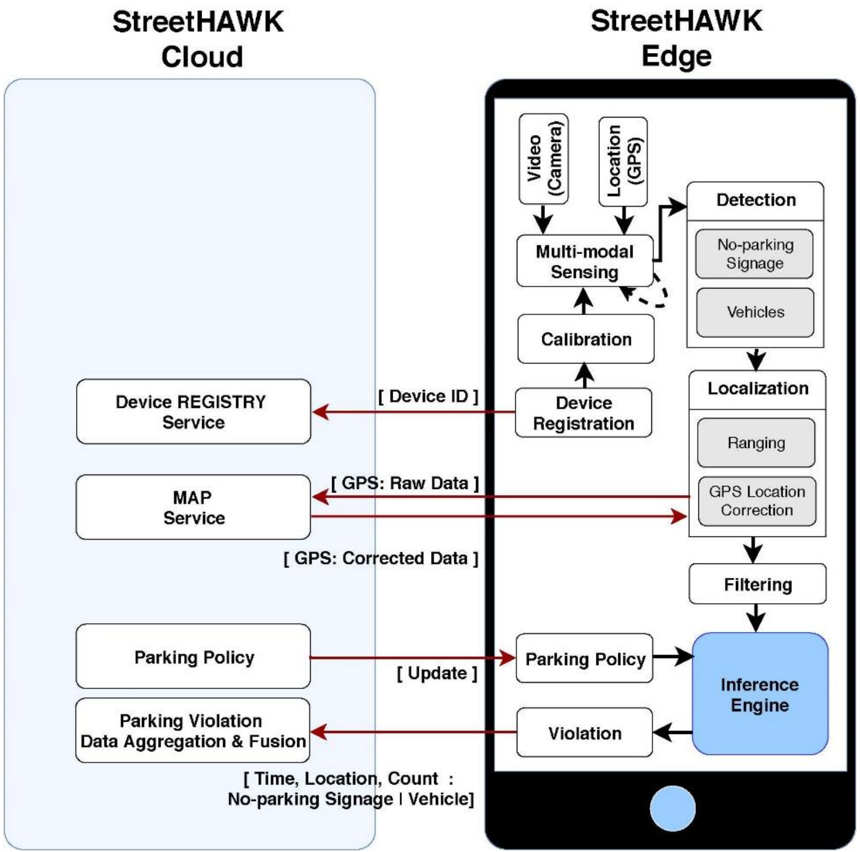
Scope of Operation

- View the roadside parking scene using the rear camera of the dashboard/windshield mounted smartphone
- Locally process the video feed in real-time to identify no-parking signages and vehicles parked in its vicinity
- Match it with parking policies to detect the location of illegally parked vehicles
- Send the status report to the cloud that can further notify various city agencies

Features

- Easy to Scale
  - City taxi networks are a widespread mode of transport | Egde based system
- Low in Cost and Deployment Complexity
  - Makes use of off-the-shelf smartphones that are already present in taxis
  - Precludes the need for additional instrumentation, either in urban environment or on the vehicle
- Real-time and Privacy Preserving
  - Performs on-board analytics on the phone itself, without sending back raw video data to the cloud
  - The status report is sent to the cloud, which consists of the location of the illegally parked vehicles
- Automated
  - On-time installation of the StreetHAWK app on the smartphone and granting it the correct system access privileges is ALL that is needed from the user point of view. The rest of the workflow is automated.
- Non-disruptive to the normal process

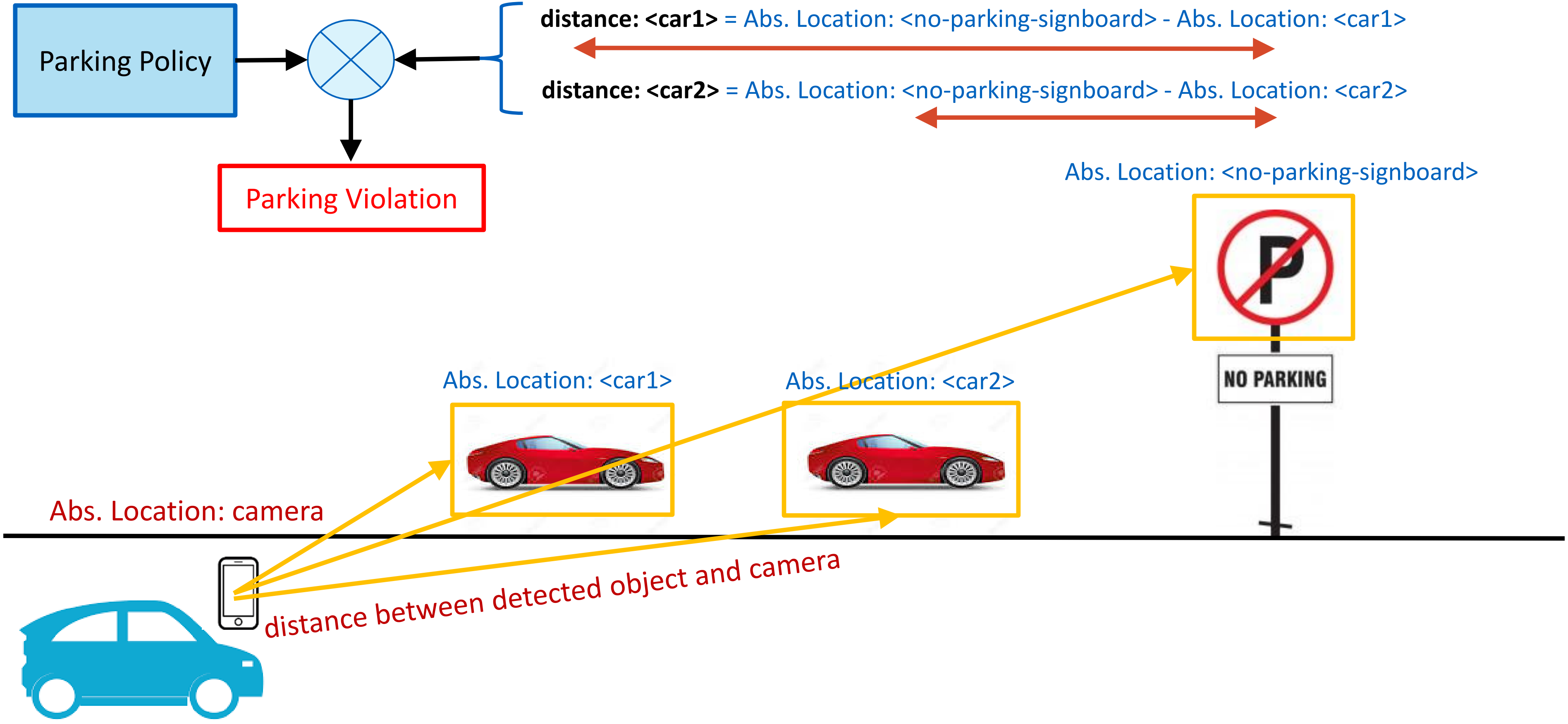
StreetHAWK: High-level Architecture



System level Challenges

- High-speed mobile vision with an inverted monitoring architecture
- "Zero" control over the set-up
- Edge platforms are constrained (in comparison to the cloud infra)
- Smartphone based edge system offer a shared working environment | cannot get greedy about platform resources

StreetHAWK: Functional Flow

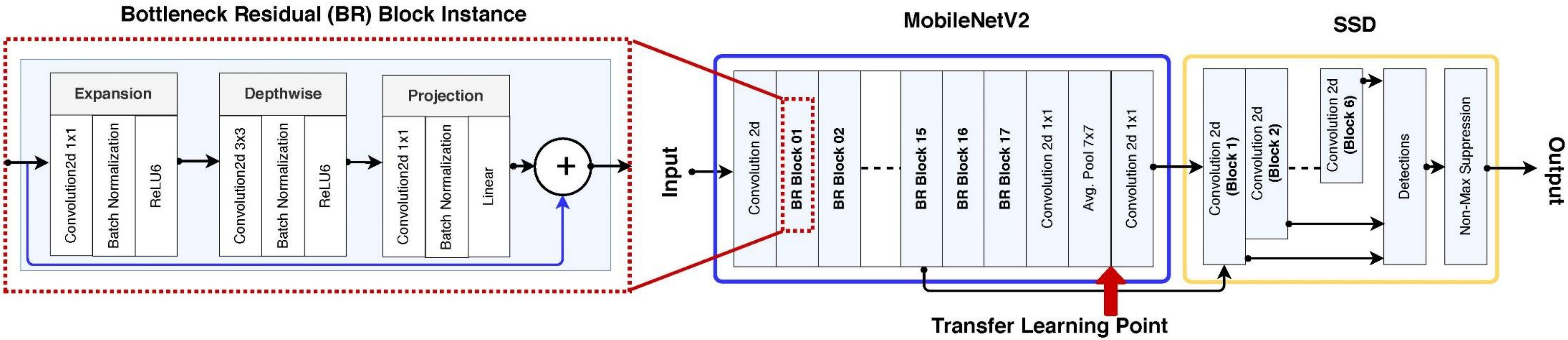


StreetHAWK: Object Detection Challenges

- **Aim**: Detect objects-of-interest (no-parking signboards | motorbikes | auto-rickshaws | cars)
- **Challenges**:
  1. Small object detection
  2. Lack of consistent pattern for visual detection (*more applicable to developing countries*)
     - Irregularities in the deployment of no-parking signboards
     - Non-standard manner of vehicle parking
  3. Unique object identification
     - Multiple detection of the same object across frames
  4. Differentiate : moving vs. parked vehicle
     - Recall: system uses a mobile setup | camera is moving | scene is moving

StreetHAWK: Object Detection Model

Convolutional detection model : SSD meta-architecture with MobileNetV2 as the feature extractor

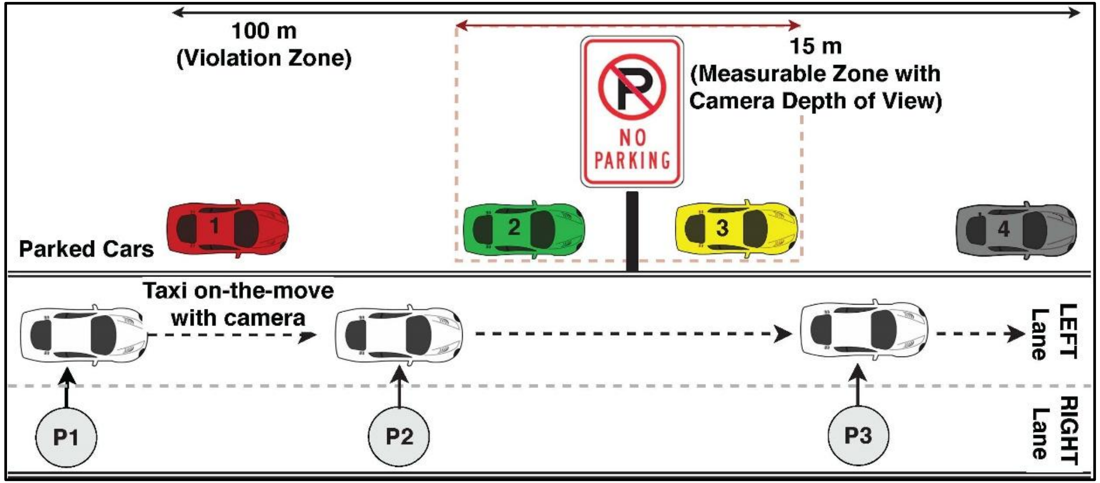


- **Approach**
  - Used a detection model pre-trained on the MS COCO dataset
  - Limited object categories to 4!
  - Derived a new model by using transfer learning and re-training with custom dataset collected under a wide range of real-world constraints and environmental conditions (with data augmentation)

Overcomes object detection challenges 1 and 2 !

StreetHAWK: Ranging Challenges

- **Aim**: Find the distance between the camera and the detected objects
- **Challenges**:
  - Since the camera is moving, there is a need for an ultra-fast visual distance measurement technique
  - Limited measurement range with a single camera system
    - The single shot detection range is limited to 15m, while the field requirement is that of 100m

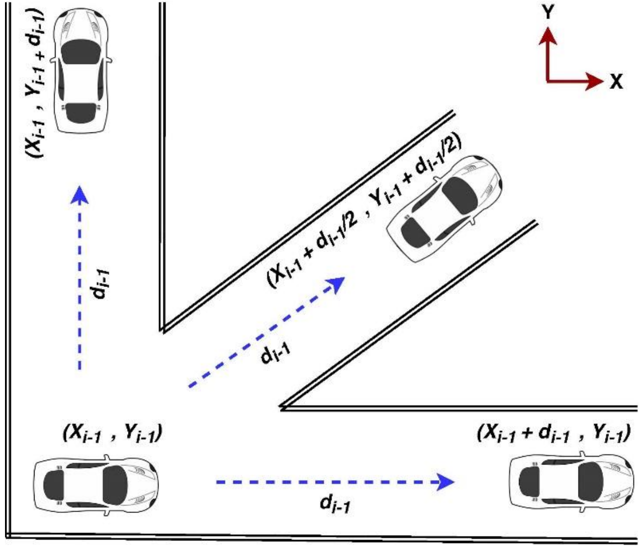


- **Approach**:
  - Used the pinhole principle to measure the distance between the camera and the detected objects
  - Used a short-term historian to successively log the details of the detected objects

StreetHAWK: Localization Challenges

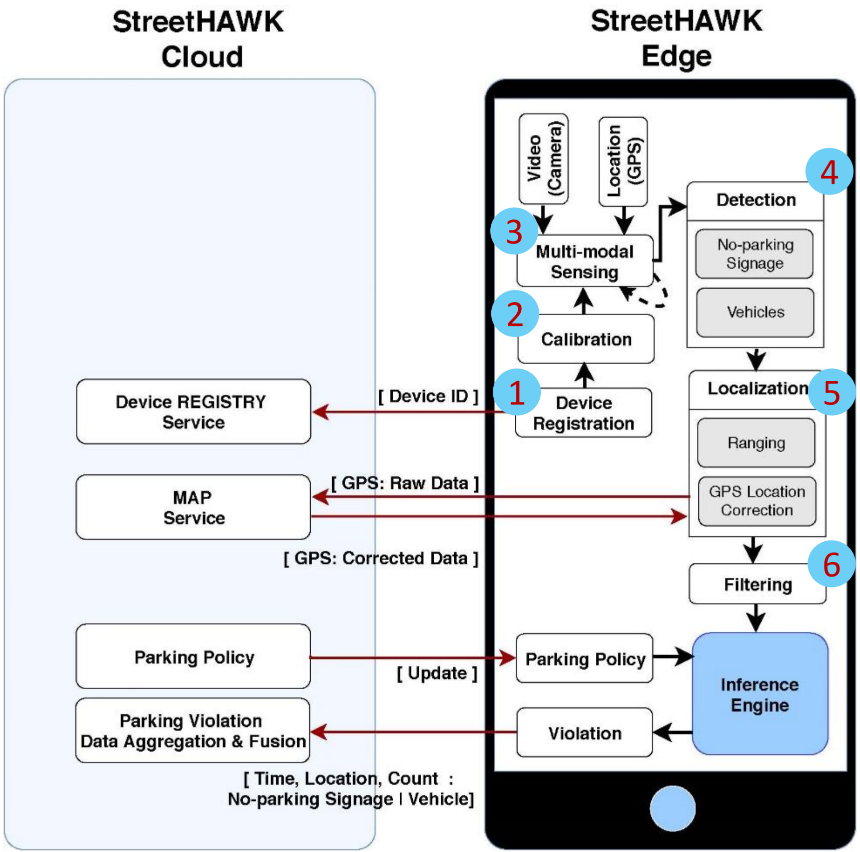
- **Aim**: Find the absolute location of the detected objects

- **Challenges**:
  - For real-time operation:
    - localization needs to be performed on the device itself
    - localization technique must be lightweight



- **Approach**:
  Find out if <x, y, or both x and y> co-ordinates of the camera location are changing
  - Wait for at least 2 consecutive object detections
  - If <x> co-ordinate is changing, then add the Distance (camera->object) to x!
  - If <y> co-ordinate is changing, then add the Distance (camera->object) to y!
  - If both <x and y> co-ordinates are changing, then add Distance (camera ->object)/2 to both <x> and <y>

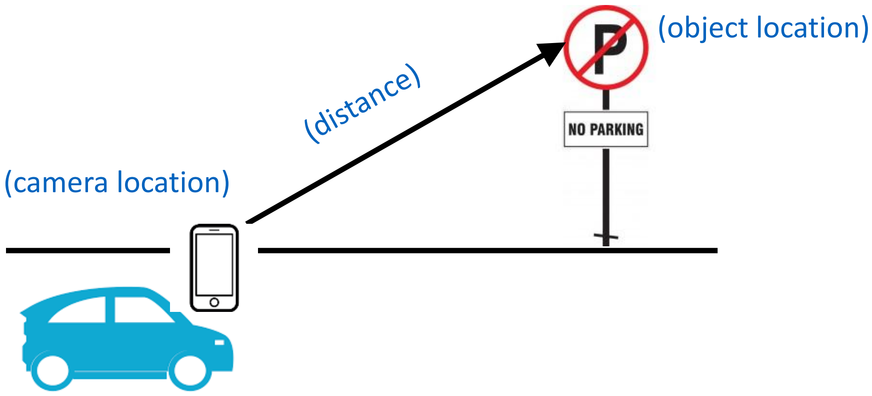
StreetHAWK: Functional Flow



Historian

Record (for EACH detected object):

- (Time) of detection
- Camera (location)
- (Distance) between camera and detected object
- Object (location)

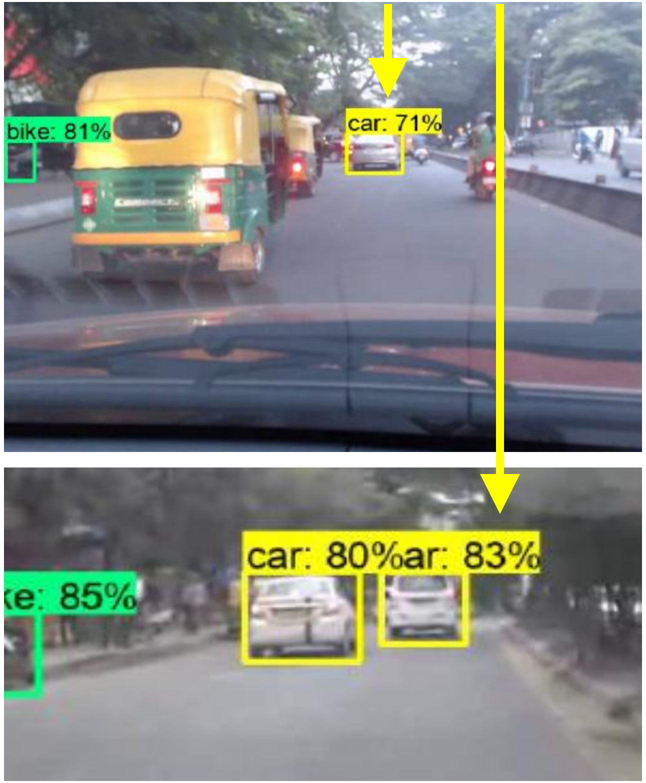


StreetHAWK: Filtering

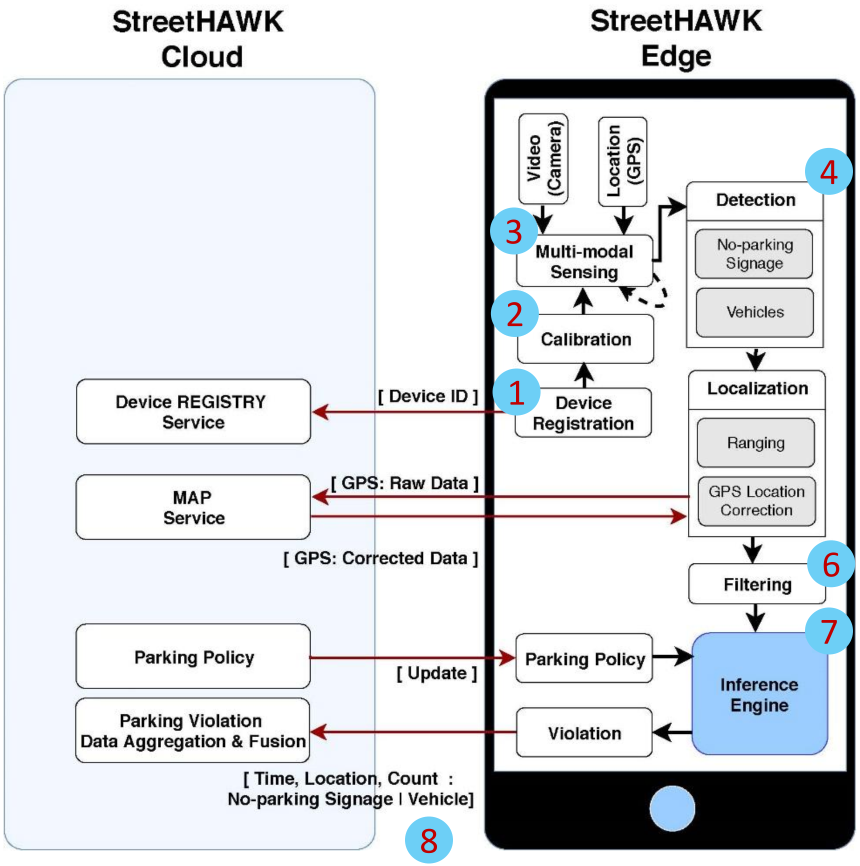
- **Aim**: remove erroneous object detections
- **Challenges**:
  1. Unique object identification
     - Multiple detection of the same object across frames
  2. Differentiate : moving vs. parked vehicle
     - Recall: system uses a mobile setup | camera is moving | scene is moving

- **Approach**:
  - Unique object identification
    - Perform sequential clustering of object locations
      - Calculate the distance between two consecutive object locations
      - If distance < 1m, then the object locations are clustered together as belonging to the same detected object
  - Differentiate between a parked and a moving vehicle
    - If the camera is moving, and as the camera gets closer to the detected vehicle:
      - If the distance between the camera and the detected vehicle decreases, then detected vehicle = PARKED
      - For anything else, the detected vehicle is moving

Moving vehicles are detected as PARKED



StreetHAWK: Functional Flow @ Inferencing



Historian
For each detected object: ▪ (Time) of detection ▪ Camera (location) ▪ (Distance) between camera and detected object ▪ Object (location)

- Trigger the inference engine when a no-parking signage is detected
- Calculate the distance of all parked vehicles to the no-parking signage (refer to HISTORIAN)
- Based on parking policy, flag VIOLATIONS!

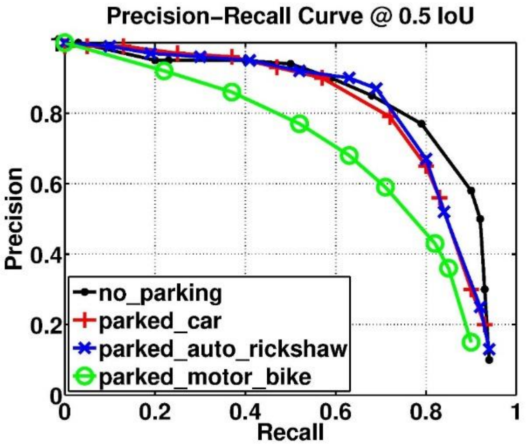
StreetHAWK: Evaluation Results

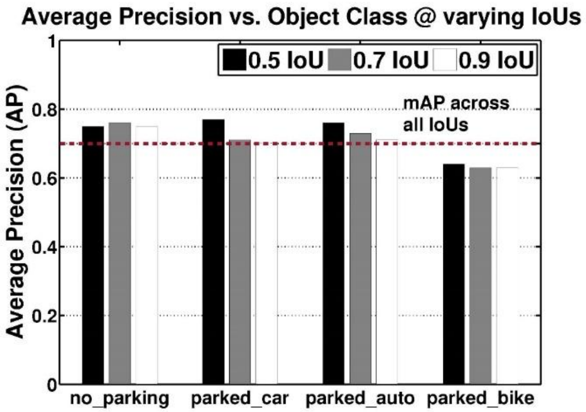
- We collected and manually labeled a dataset extracted from over 50 hours of citywide driving video feeds that spanned over 4 months
- We performed on-the-road experiments that spanned close to 500 km

StreetHAWK: Object Detection Examples

StreetHAWK: Object Detection Performance

PASCAL VOC Evaluation Benchmarks



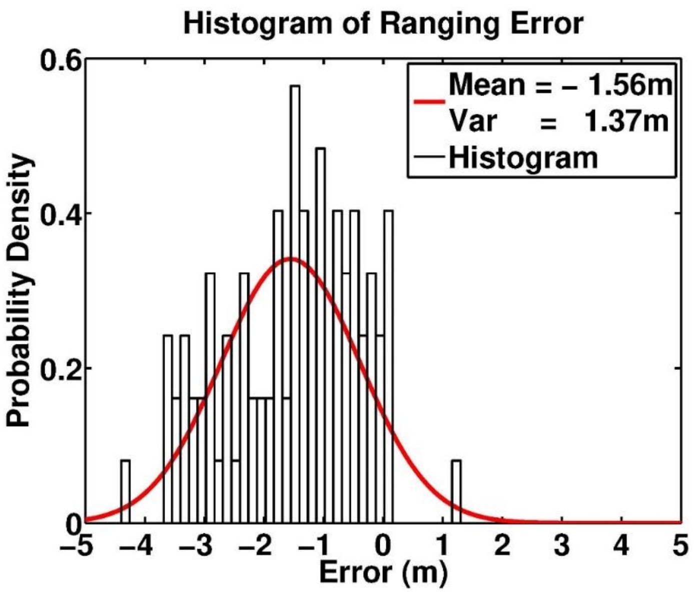


MS COCO Evaluation Benchmarks

Method	Category	mAP, IoU : (0.50 - 0.95)	(0.50)	(0.75)	mAP, Area : (S)	(M)	(L)	mAR, maxDets : (01)	(10)	(100)	mAR, Area : (S)	(M)	(L)
MobileNetV1 + Faster RCNN	Embedded	0.17	0.41	0.10	0.01	0.17	0.28	0.21	0.30	0.31	0.08	0.32	0.51
MobileNetV1 + SSD300	Embedded	0.22	0.48	0.18	0.05	0.24	0.29	0.28	0.34	0.34	0.11	0.34	0.45
VGCNet + SSD300	Non-Embedded	0.23	0.41	0.23	0.05	0.23	0.41	0.23	0.33	0.35	0.10	0.38	0.57
MobileNetV2 + SSD300*	Embedded	**0.26**	**0.55**	**0.21**	**0.07**	**0.28**	**0.37**	**0.32**	**0.42**	**0.43**	**0.20**	**0.45**	**0.58**

StreetHAWK: Ranging Performance

Static Ranging



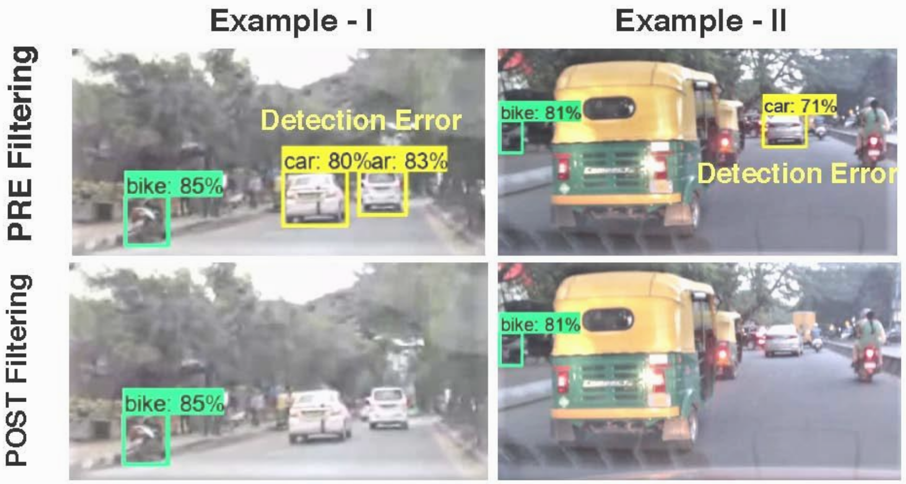
Ranging error under stationary condition, where it is observed to be less than 4 m.

Mobile Ranging

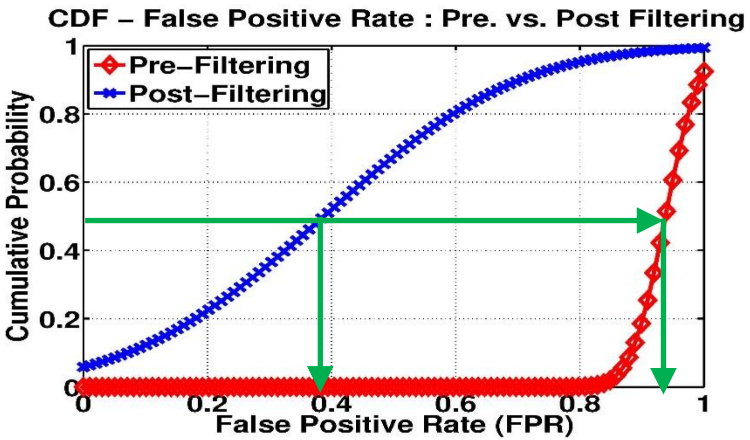
Distance (m)	Error (m), Angle :		
	$05^O \rightarrow 20^O$	$20^O \rightarrow 30^O$	$30^O \rightarrow 60^O$
$01 \rightarrow 05$	–	–	(-) 2-3
$05 \rightarrow 10$	–	(-) 3-5	–
$10 \rightarrow 15$	(-) 3-5	–	–

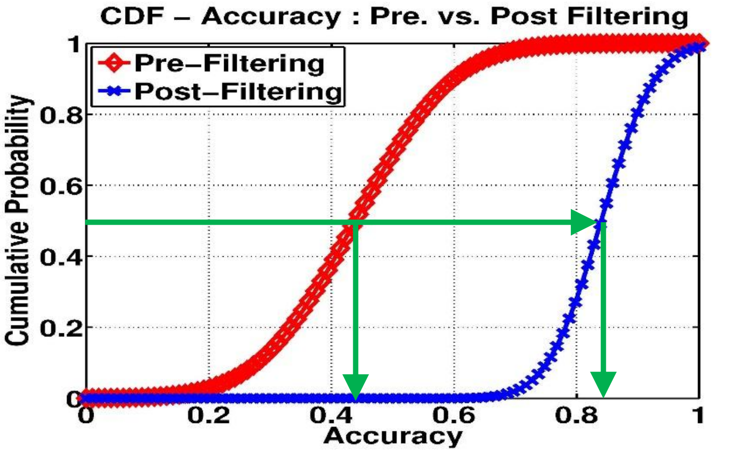
Ranging error under mobile condition; where the maximum error is observed to be less than 5m across all cases.

StreetHAWK: Filtering Performance



Representative scenarios where moving vehicles are incorrectly detected as parked, and removal of those errors post filtering.



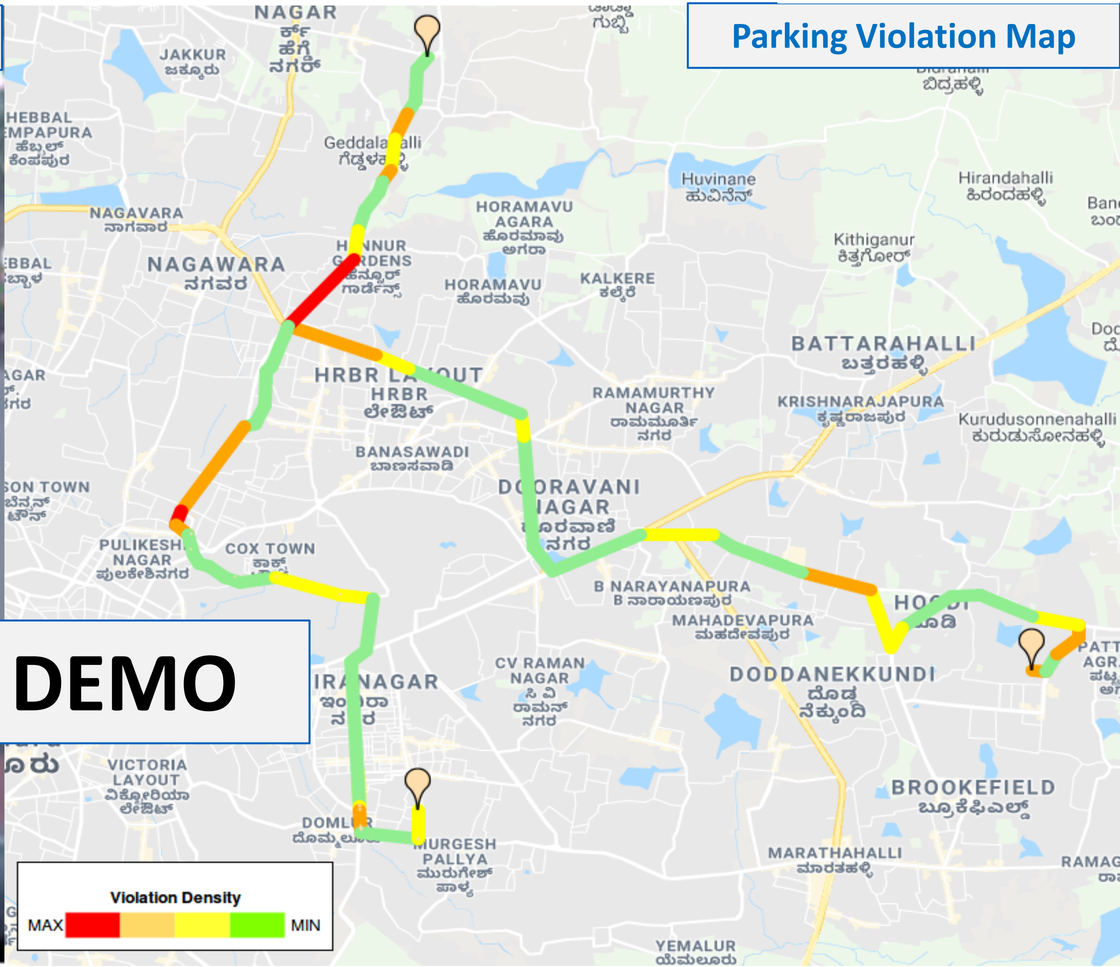


Filtering techniques significantly decrease the false positive rate; as a result of which, the end-to-end system accuracy increases by a factor of two.

StreetHAWK: On-the-road Trial

StreetHAWK UI

Parking Violation Map



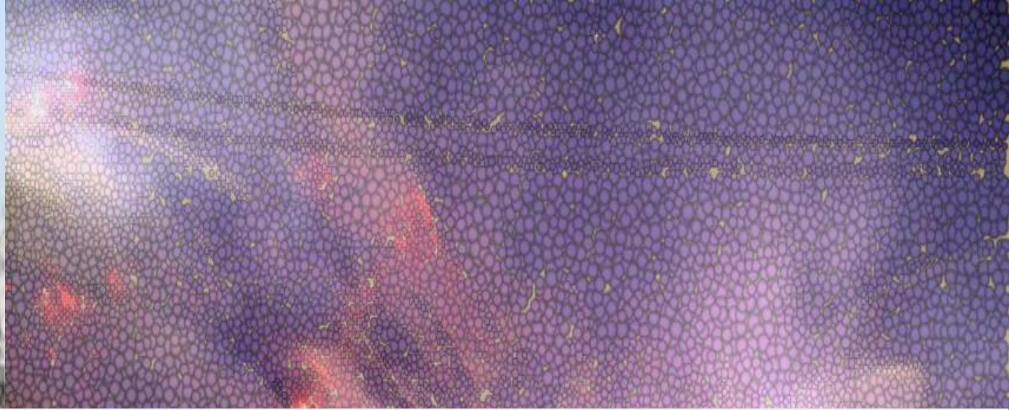
DEMO

StreetHAWK: Summary

- With an end-to-end design and implementation of StreetHAWK for parking violation monitoring, we demonstrate the feasibility of a COTS smartphone based edge system!
  - combines a single camera visual sensing mode with edge-compatible machine learning and analytic models.

- We account for the edge platform constraints and propose lightweight methods for detecting parking violations and measuring the violation span/density.
  - We use a deep neural network (DNN) based convolution detection model, and address the model limitation of identifying small objects in a wide variety of real-world conditions by extensive training and parameter tuning.
  - We use the visual ranging capability of a single camera phone system to measure the violation span, but enhance it with a short-term historian and GPS to extend the system range from 15m to the prescribed 100 m.
  - At the overall system level, we make use of the mobility of the camera unit and multi-modal sensing clues to filter out erroneous violation instances.

- System performance
  - three times better accuracy in detecting small sized objects than other state-of-the-art embedded models
  - worst case ranging error of less than 5 m
  - operates at a speed of 5 frames per second (FPS) on typical mid-range Android smartphones
  - identifies, on an average, 80% of parking violations compared to a perfect record of manual approaches

Use-case: Drone aided Aerial Search & Emergency Response

(Published in IEEE ICRA'18)


- P. Misra, A. Anil Kumar, P. Mohapatra, and Balamuralidhar P., "DroneEARS: Robust Acoustic Source Localization with Aerial Drones" *IEEE International Conference on Robotics and Automation (ICRA),* pages 80-85, 2018. IEEE.

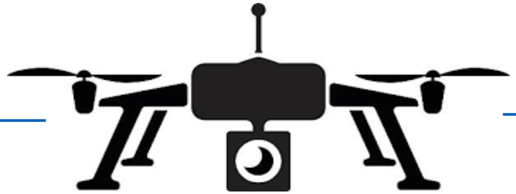
Limitation: Aerial Drones for Search operations?

Challenge - 1

Unusable under Camera **Obstruction**

Challenge - 4

Unusable under **Non Visual Clues**



Challenge - 2

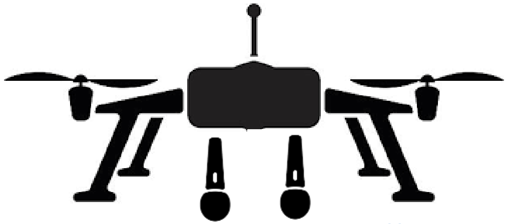
Unusable under **Occlusion**

Aerial Drones with “EYES”

Challenge - 3

Unusable under **Low Visibility**

Proposal: Drone-EARS



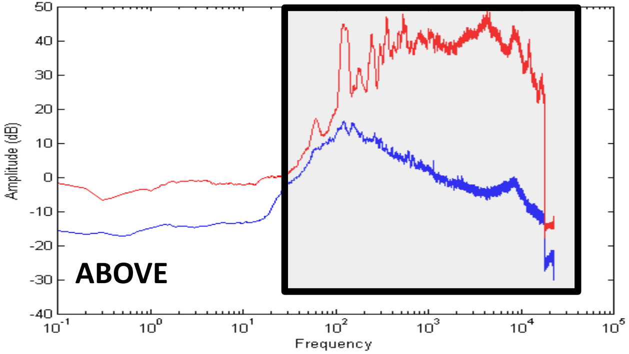
Aerial Drones with "EARS"

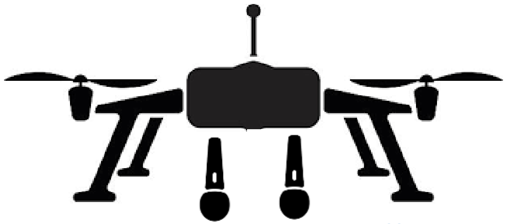
MIT TR35 India 2017

https://www.youtube.com/watch?v=gDAXAFbCVOQ

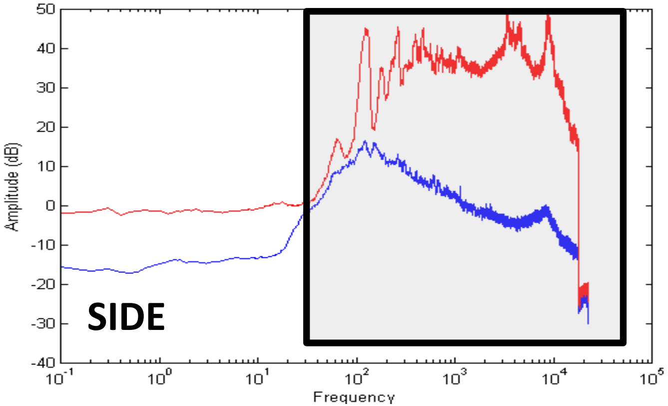
The Road Blocker?

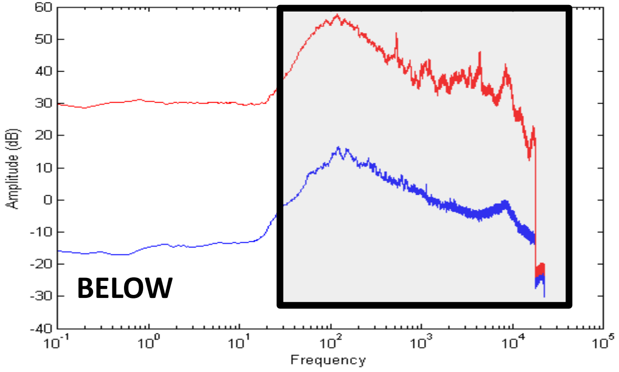
STRONG “EGO-NOISE”: Mechanical and Wind Noise





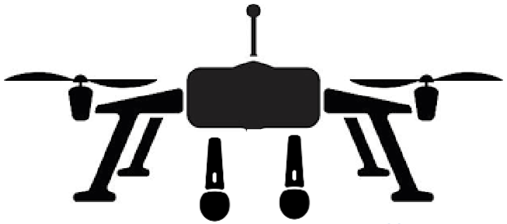
Aerial Drones with “EARS”





The Mitigation Strategy

Introduce **Large** **SPATIAL-TEMPORAL** diversity to boost the Received SNR

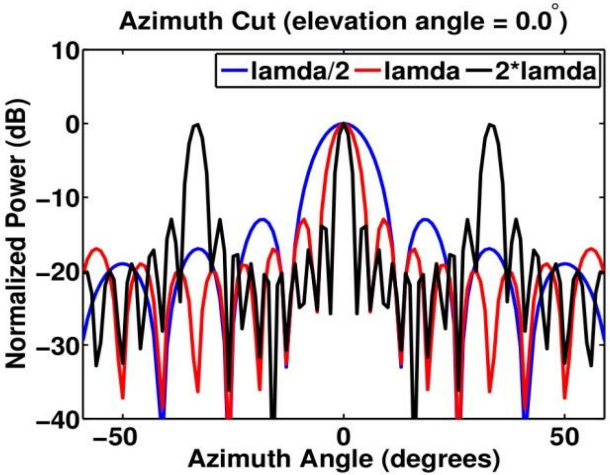


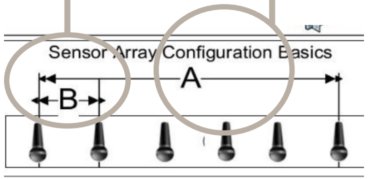
Aerial Drones with "EARS"

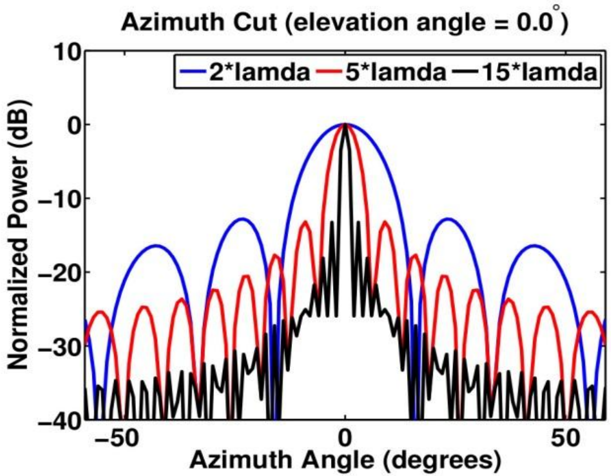
Challenge in the Mitigation Strategy

Large Microphone Arrays **vs. Micro** Aerial Drone **Platform** Constraints

A Glaring Contradiction !!!







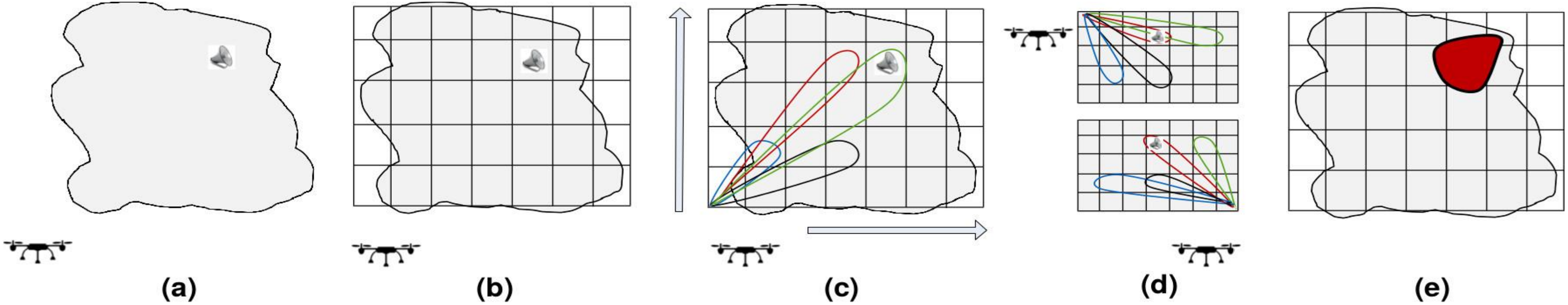
- **Larger A** -> narrower beam | can better resolve sound sources
- **B < (wavelength / 2)** -> avoids undesirable grating lobes that results due to spatial aliasing

Microphone Array Design Rationale:

Sparse array with **large** aperture length (A), but using **fewer** array elements with **larger** B

Illustration of the Solution - I

Mobility-aided Beamforming for Sparse Microphone Array Models



Note: the speaker symbol in (a)-(d) denotes the original location of the sound source, and the red blob in (e) represents its estimated location.

Intra-band Fusion (Step-1) : For a given measurement location ***l***. the power corresponding to each cell location across N_s sub-bands is fused as :

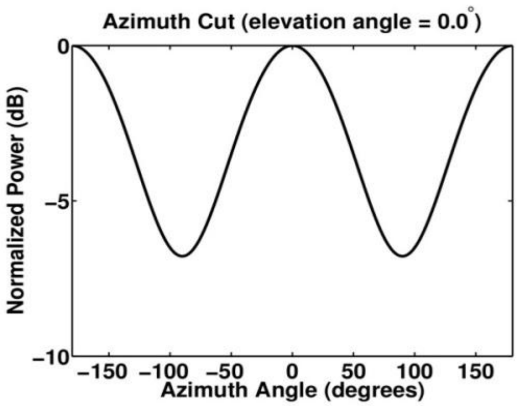
$$\Delta_{l,\tilde{\mathbf{p}}_k} = \prod_{i=1}^{N_s} \Gamma^{(i)}_{l,\tilde{\mathbf{p}}_k}$$

where $\Gamma^{(i)}_{l,\tilde{\mathbf{p}}_k}$ denotes the i^{th} sub-band power corresponding to the cell location $\tilde{\mathbf{p}}_k$, and the measurement location ***l*** (estimated using the standard Capon's beamformer).

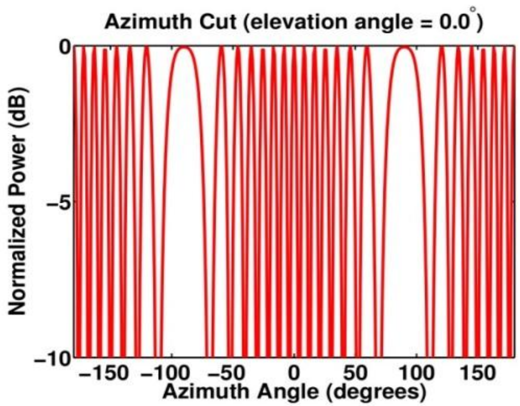
Inter-measurement Fusion (Step-2) : The power across the ***L*** measurement locations are fused, and the final aggregated power $\tilde{\mathbf{p}}_k$ corresponding to the location is estimated as :

$$W_{\tilde{\mathbf{p}}_k} = \sum_{l=1}^{L} \Delta_{l,\tilde{\mathbf{p}}_k}$$

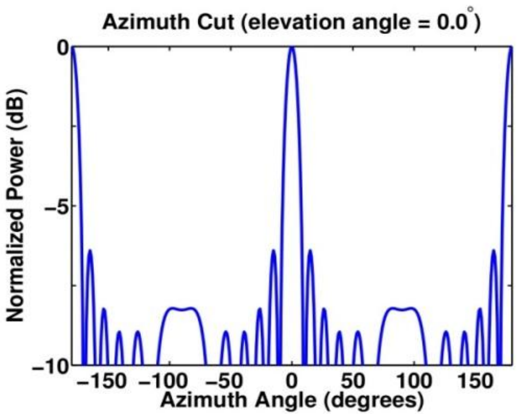
Illustration of the Solution - II



Beampattern of Lower Frequency : 500 Hz



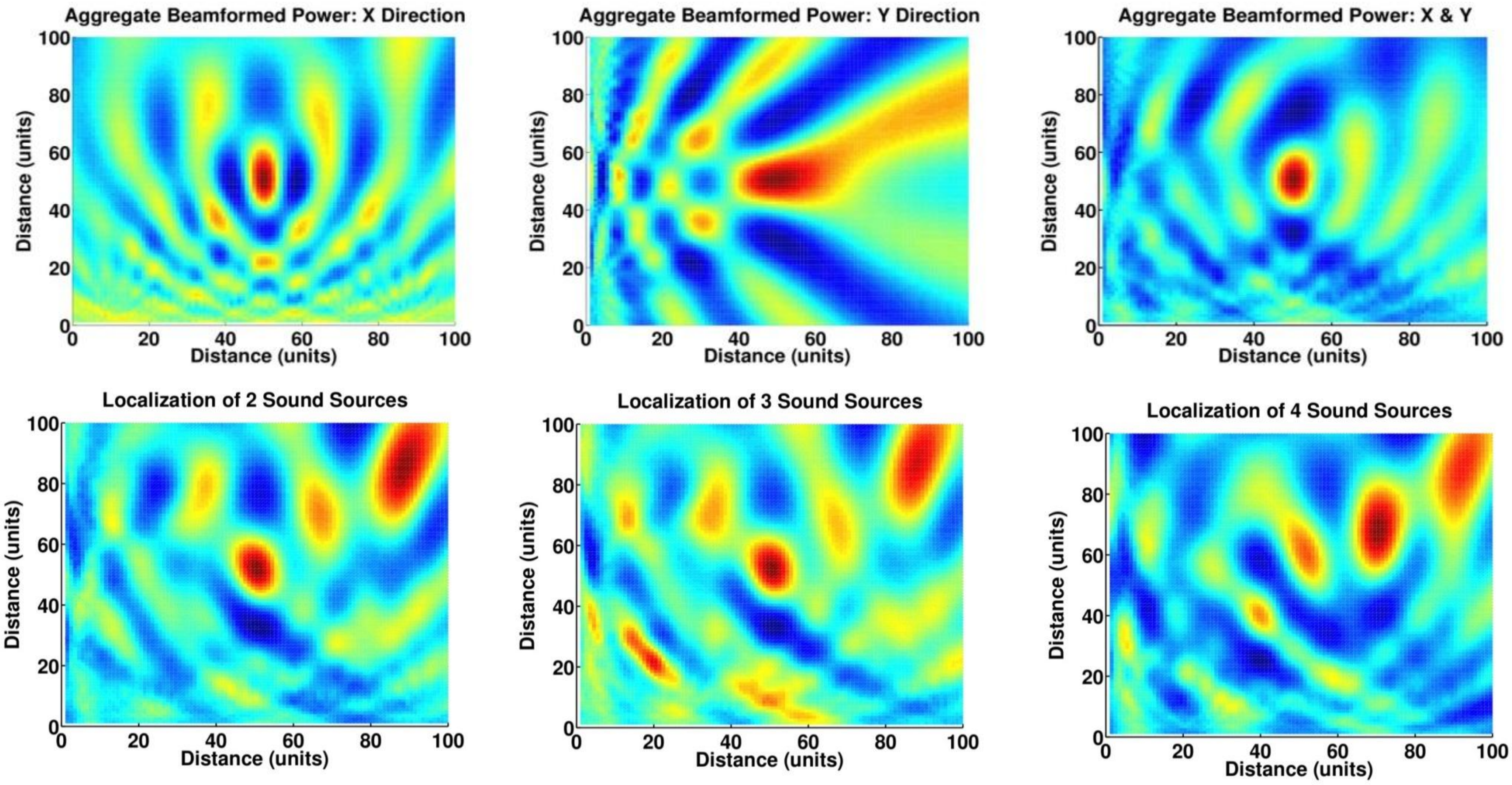
Beampattern of Lower Frequency : 5 kHz



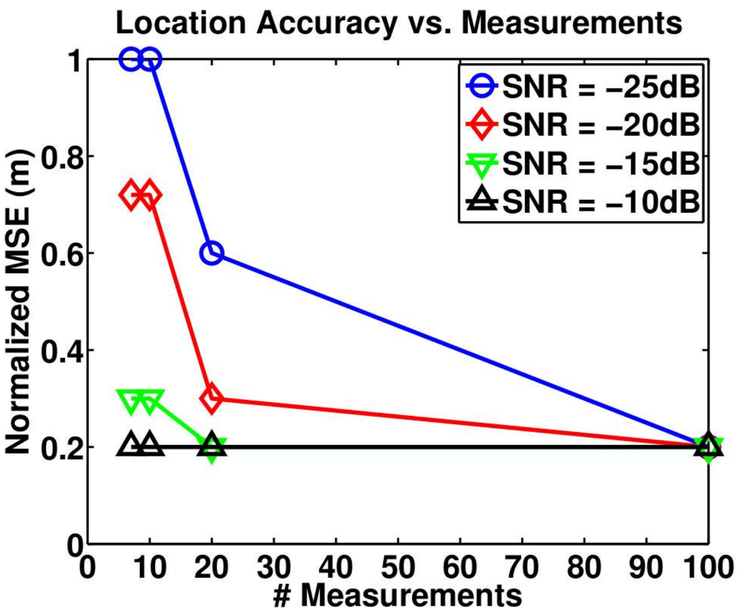
Intra-band Beam Fusion

- [LEFT] shows the beam corresponding to 500 Hz (single beam at 0^o, but of wider width).
- [CENTRE] shows the beam corresponding to 5 kHz (narrow beam, but with grating lobes at 180^o that spreads into -90^o and $+90^o$).
- [RIGHT] The **intra-band fusion** leads to a relatively **narrow beam, but with significant reduction in grating lobes.**

Result - I


Aggregate Beamformed Power: X Direction
Aggregate Beamformed Power: Y Direction
Aggregate Beamformed Power: X & Y
Localization of 2 Sound Sources
Localization of 3 Sound Sources
Localization of 4 Sound Sources
Distance (units)
0
20
40
60
80
100

Result - II



- Received signals of **very low SNR need more number of mobile measurements** to achieve an average measure of localization accuracy compared to ones with higher SNR levels.
  - This effect is very profound at the received SNR level of −25 dB, where at least 20 such spatial measurements are required to reduce the MSE by 50%.
  - However, the number of measurements needed for signals that are received with higher SNR (−20 dB and above) levels to obtain a good location estimate is quite small.

Computationally Exhaustive !!!

Drone-EARS: Summary

Potential Impact

- Enable acoustic scene analysis
- Enable multi-modal sensing on aerial drones, something that is yet not available in COTS form

Business Domains

- Industry 4.0

Application Use Cases

- Humanitarian - Emergency Response, Search & Rescue

Features

- Acoustic source localization at **low SNR** levels

Use-case : Establishing the Provenance of Diseases*

*** This work is done PRIOR to joining TCS Research.**

Tracking & Monitoring Megabats: Flying Foxes

Disease Vector

- Hendra Virus
- Ebola
- SARS-like Coronavirus

Agricultural pests

- Crop Damage

Interaction

- With other flying foxes and animals

Behavior

- Not well understood
- Threatened species

<u>P. Misra</u>, W. Hu, Y. Jin, J. Liu, A. S. de Paula, N. Wirstrom, and T. Voigt, “Energy Efficient GPS Acquisition via Sparse-GPS”, In *Proceedings of the 13th ACM/IEEE Conference on Information Processing in Sensor networks,* IPSN’14, pages 155-166, 2014. IEEE.

Sparse-GPS: Summary

Highlights

- We introduced Sparse-GPS, a new computing framework for energy efficient GPS acquisition via sparse approximation
  - Proposed: A new dictionary that combines the information sparsity along all search dimensions

- Analyzed the dependency of the received SNR and satellite acquisition count on data length
  - Showed: Using 10-20 ms over 2ms of data, there is a high probability of acquiring 50% additional satellites with both the conventional and S-GPS

- Demonstrated the GPS acquisition capability and energy gains by empirical evaluations on real GPS signals.
  - Showed: S-GPS is 2 times more energy efficient than offloading uncompressed data
  - Showed: S-GPS is 5-10 times more energy efficient than standalone GPS
  - Showed: S-GPS has a median positioning error of 40m

Summary of the Talk

- Mobile Sensing & IoT at Scale.
- Covered three case studies:
  - Visual sensing with mobile phones (for monitoring parking violations)
  - Acoustic sensing with drones (for emergency response),
  - RF sensing (GPS) with sensor nodes (for 'ultra' long-term tracking)
- Discussed different mobile sensing system designs that piggyback on the platform mobility to improve the sensing and interpretation performance, while adhering to the system limitations and application scope.