

# Object Tracking

**Samreen Anjum & Danna Gurari**

The University of Texas at Austin

Fall 2019



# Review

- Last week:
  - Video classification and localization applications
  - Evaluating video classification and localization
  - Crowdsourcing video classification and localization
- Assignments (Class Website & Canvas)
  - Project proposal due yesterday
  - Project outline due in two weeks
- Questions?

# Today's Topics

- Object tracking applications
- Evaluating object tracking methods
- Crowdsourcing object tracking strategies
- Object tracking datasets & challenges
- Lab: video annotation & creating figures/tables in latex

# Today's Topics

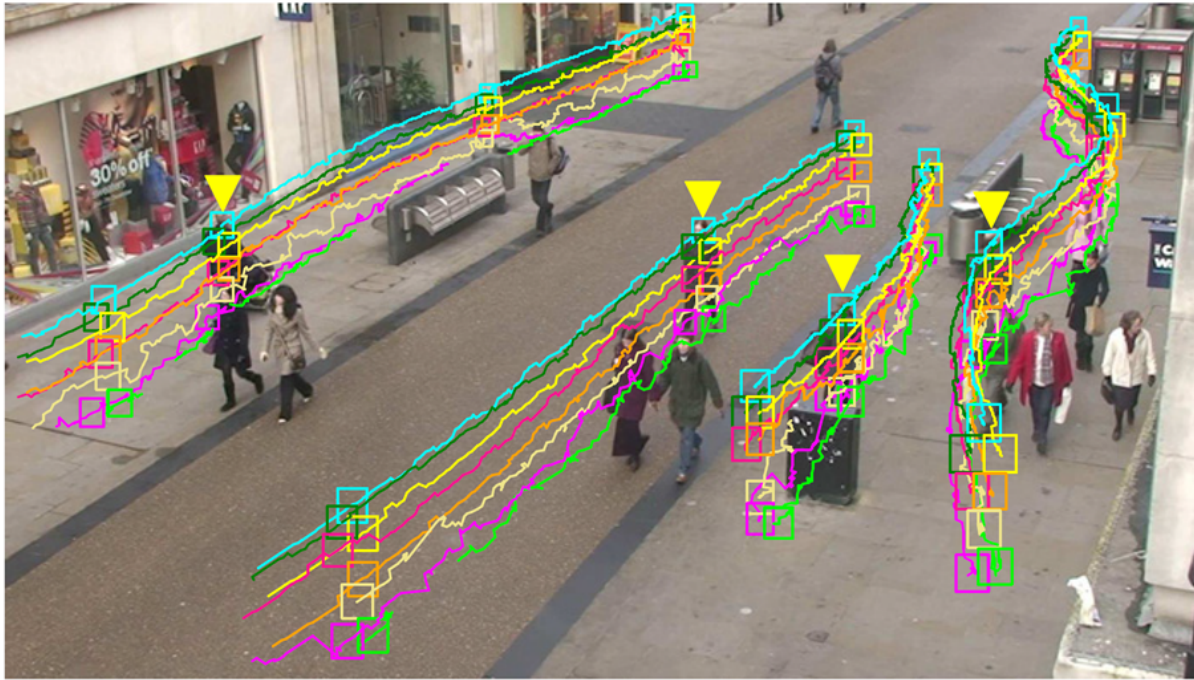
- Object tracking applications
- Evaluating object tracking methods
- Crowdsourcing object tracking strategies
- Object tracking datasets & challenges
- Lab: video annotation & creating figures/tables in latex



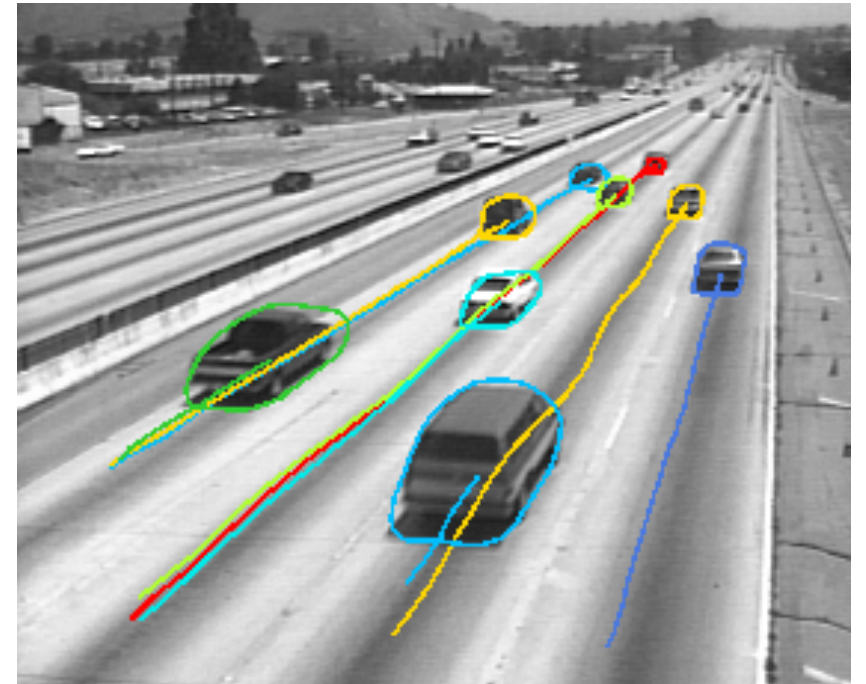
# Definition

- Each object in a video is localized (object detection) and followed as it moves over time (tracking)

# Application: Surveillance



People Tracking

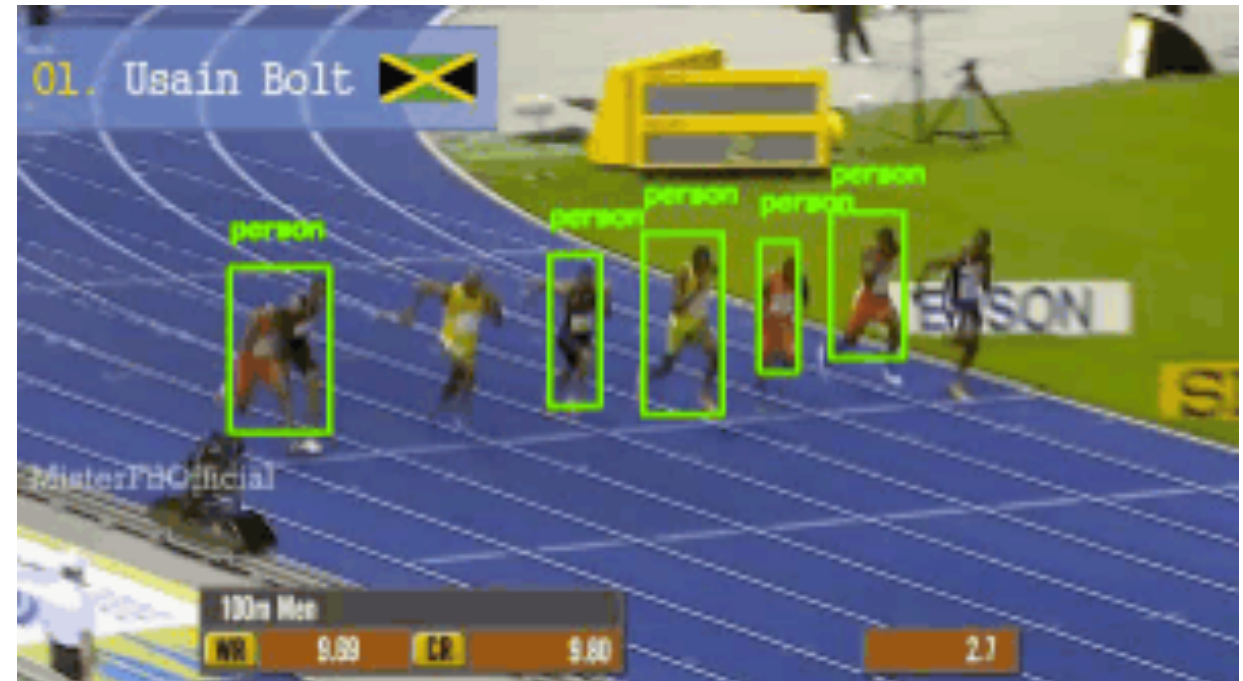


Traffic Control

<http://www.vision.caltech.edu/koller/MOU-83.html>

<https://heartbeat.fritz.ai/the-5-computer-vision-techniques-that-will-change-how-you-see-the-world-1ee19334354b>

# Application: Sports Analysis





# Application: Performance Analytics

Calculate Bat speed from video!



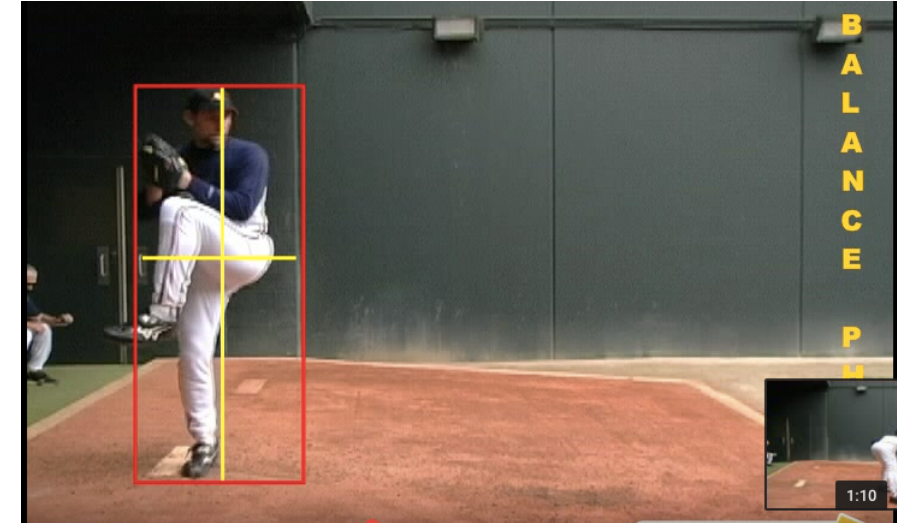
NEW! Track Bowling Ball Path!



Works great for putting!



<http://www.motionprosoftware.com/>



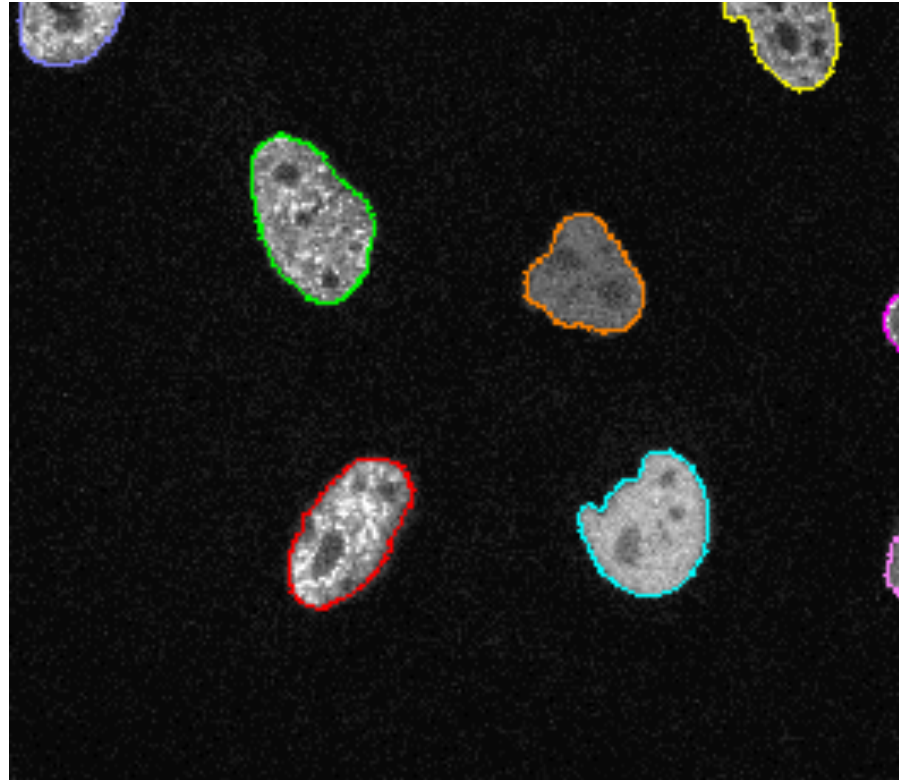
[Watch: Dartfish Analysis](#)

[Watch: Pitch Analysis](#)

# Application: Military Defense



# Application: Cell Tracking & Mitosis Detection



# Application: Self-driving Cars





# Application: People Counting



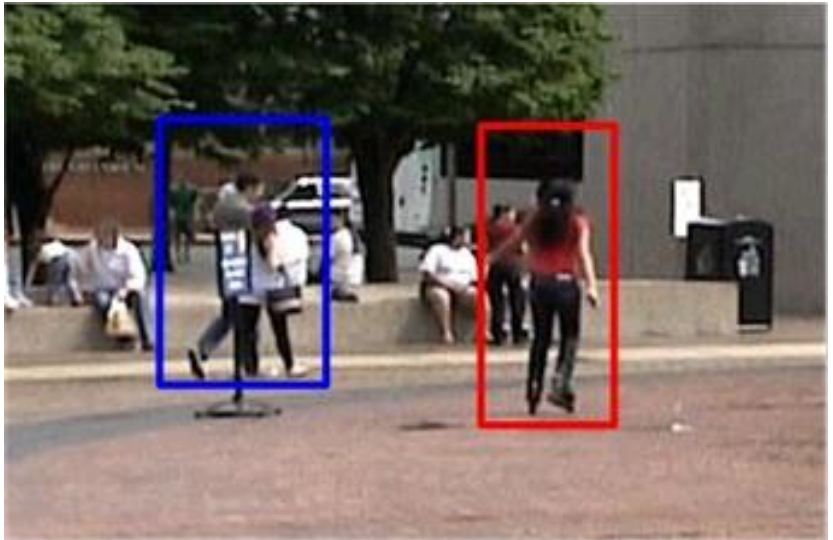


# Application: Human Computer Interaction



Roboceptionist

# Application: Activity Recognition



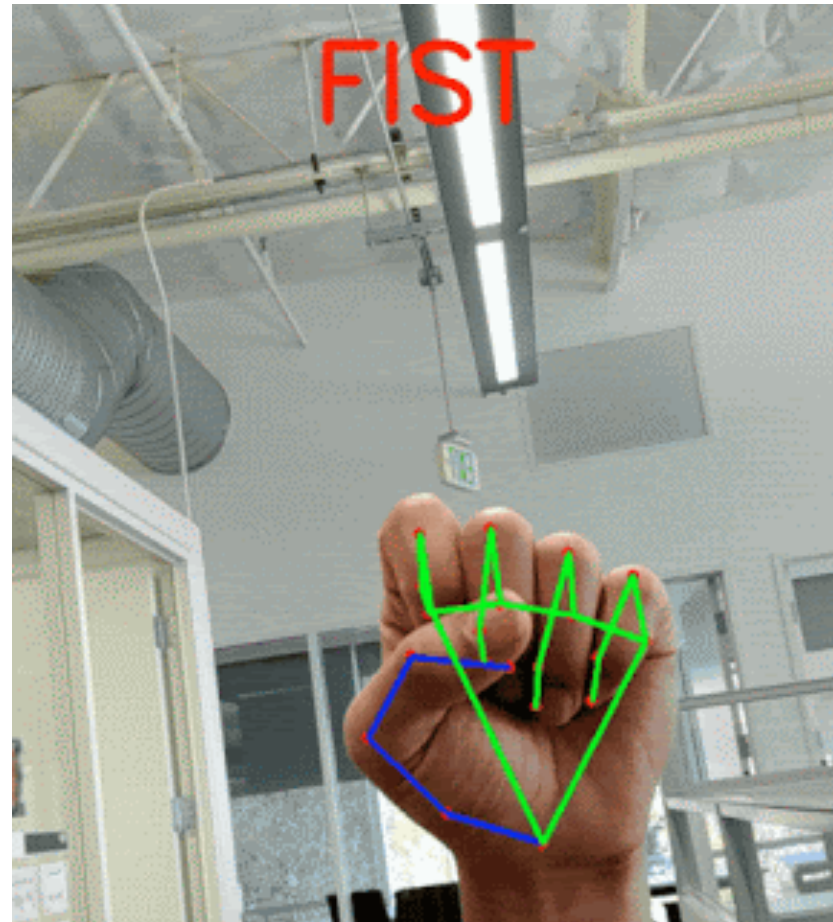
walk

skate

<https://www.bu.edu/ids/research-projects/action-recognition/>

[https://www.crcv.ucf.edu/projects/realistic\\_action\\_recognition/](https://www.crcv.ucf.edu/projects/realistic_action_recognition/)

# Application: Sign Language Recognition





# Application: Marketing



# Application: Augmented Reality



<https://www.geekwire.com/2017/augmented-reality-shopping-phone-patent-hints-amazons-aspirations/>

<https://virtualrealitypop.com/object-recognition-in-augmented-reality-8f7f17127a7a>

# Applications

What other applications can you think of where object tracking could be useful?

# Applications: Many More

- Robotics
- Space and satellite
- Industrial production
- Ecological monitoring
- Healthcare
- Anomaly detection

# Today's Topics

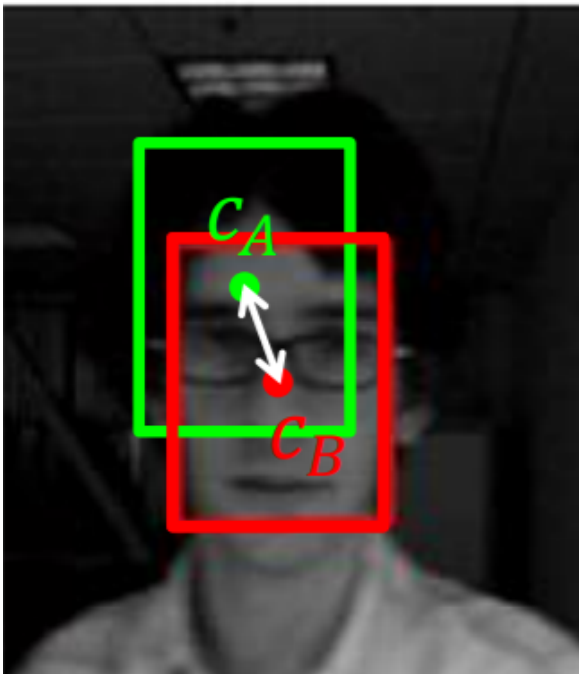
- Object tracking applications
- **Evaluating object tracking methods**
- Crowdsourcing object tracking strategies
- Object tracking datasets & challenges
- Lab: video annotation & creating figures/tables in latex



# Evaluation Metrics - I

## Precision

Distance between the centers of bounding boxes for each frame



A = Ground Truth

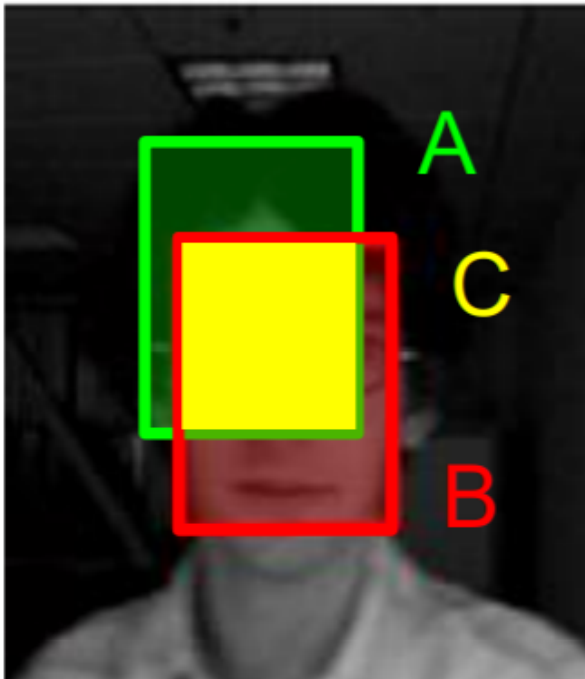
B = Predicted Track

$$p = \|c_A - c_B\|$$

# Evaluation Metrics - II

## Overlap

Intersection over Union (IoU) of the bounding boxes for each frame

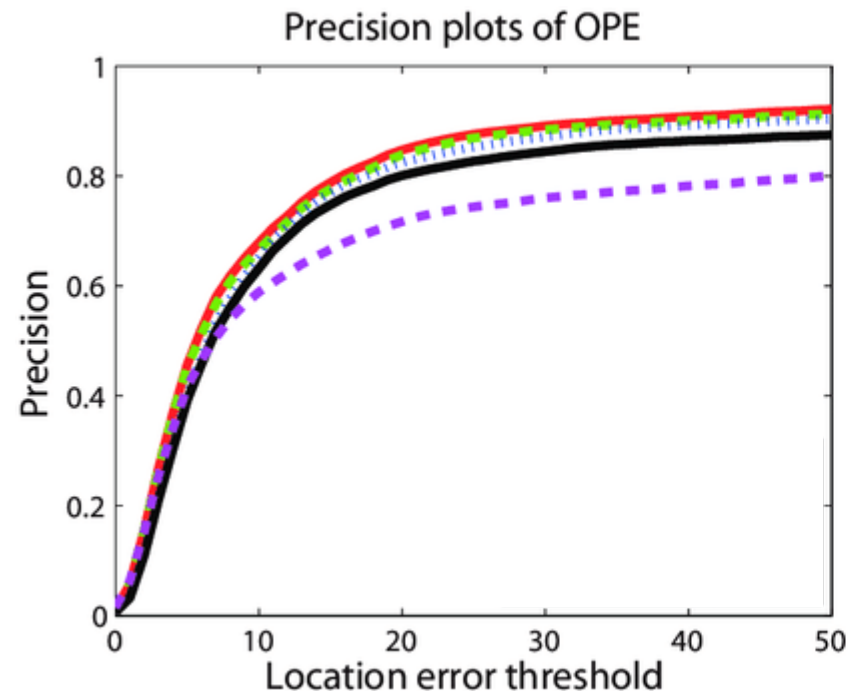


A = Ground Truth  
B = Predicted Track  
C = Intersection

$$IoU = \frac{C}{A + B - C}$$

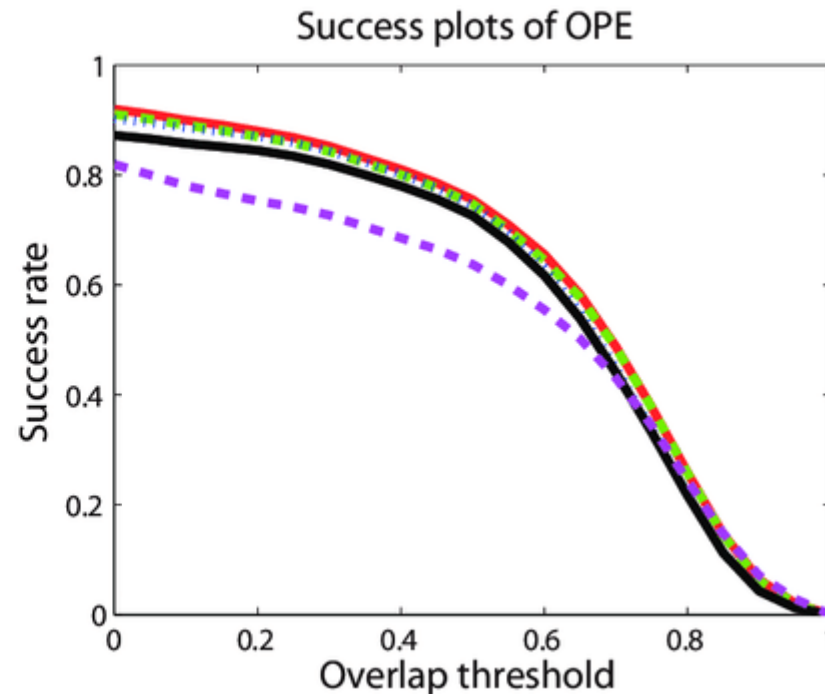
# Visualization of Evaluation Metrics - I

- **Precision plot:** Percentage of frames with predicted location within a given threshold distance of ground truth (for e.g., 20 pixels)



# Visualization of Evaluation Metrics - II

- **Overlap plot:** Percentage of frames where the overlap between the bounding boxes is larger than a given threshold (for e.g., 0.5)



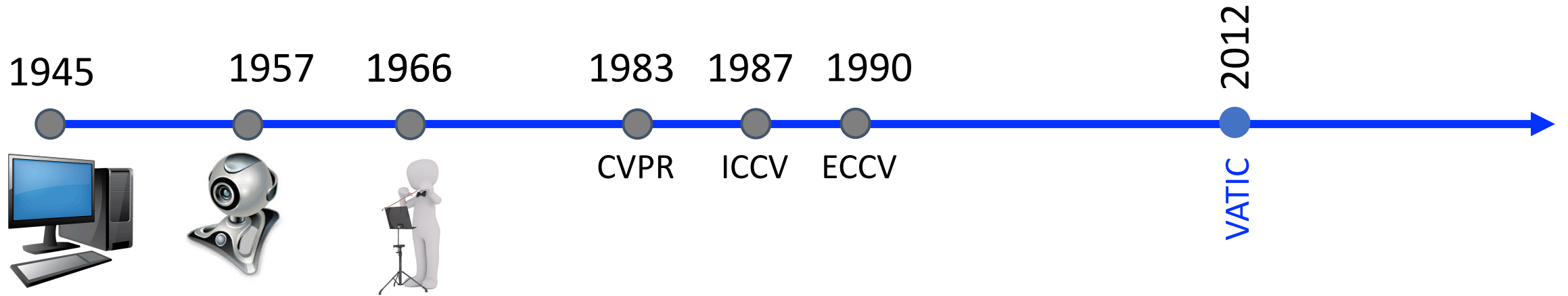
# Evaluation References

- Gao, C., Shi, H., Yu, J. G., & Sang, N. (2016). Enhancement of elda tracker based on cnn features and adaptive model update. *Sensors*, *16*(4), 545
- Wu, Y., Lim, J., & Yang, M. H. (2013). Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2411-2418)

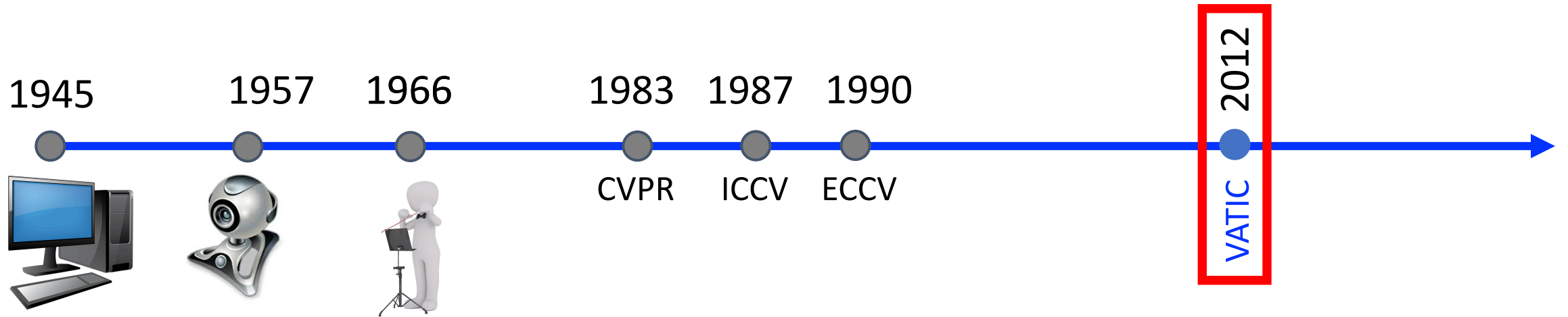
# Today's Topics

- Object tracking applications
- Evaluating object tracking methods
- **Crowdsourcing object tracking strategies**
- Object tracking datasets & challenges
- Lab: video annotation & creating figures/tables in latex

# Object Tracking Approaches



# Object Tracking Approaches: VATIC





# Object Tracking Approaches: VATIC

Annotate every object, even stationary and obstructed objects, for the entire video. Instructions + New Object

Car 12  
 Outside of view frame  
 Occluded or obstructed  
 Parked  
 Driving  
 Reversing

Person 11  
 Outside of view frame  
 Occluded or obstructed  
 Walking  
 Running  
 Standing

Car 10  
 Outside of view frame  
 Occluded or obstructed  
 Parked  
 Driving  
 Reversing

Car 9  
 Outside of view frame

Car 12 Driving  
Car 10 Reversing  
Car 1 Parked  
Car 2 Parked  
Car 3 Parked  
Car 9 Parked  
Car 8 Parked  
Car 7 Parked  
Car 6 Parked  
Car 5 Parked  
Car 4 Parked  
Person 11 Walking

◀ Rewind ▶ Play

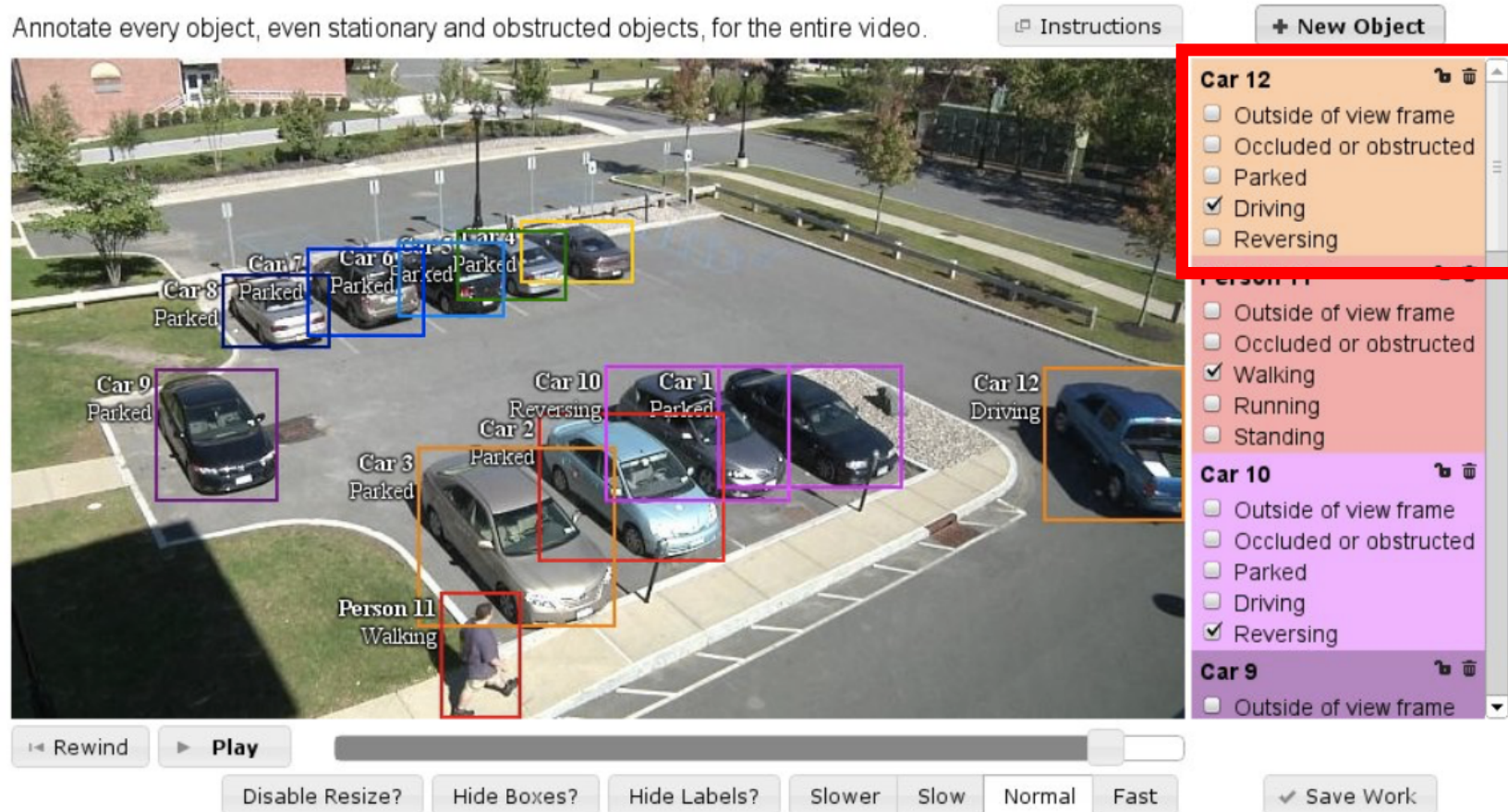
Disable Resize? Hide Boxes? Hide Labels? Slower Slow Normal Fast

✓ Save Work

Demo: <https://www.youtube.com/watch?v=ljl5pAowACc>

Carl Vondrick, Donald Patterson, and Deva Ramanan. Efficiently Scaling Up Crowdsourced Video Annotation: A Set of Best Practices for High Quality, Economical Video Labeling. IJCV 2012.

# Object Tracking Approaches: VATIC



Metadata about each object: e.g., activity, attributes, etc.

Carl Vondrick, Donald Patterson, and Deva Ramanan. Efficiently Scaling Up Crowdsourced Video Annotation: A Set of Best Practices for High Quality, Economical Video Labeling. IJCV 2012.



# VATIC Design

- Goal: user studies to establish best-practices
- Data: 3 videos



1. Scripted: large objects, linear motion



2. Basketball: small object occluded frequently, nonlinear motion



3. VIRAT: slow moving/stationary cars, linear motion

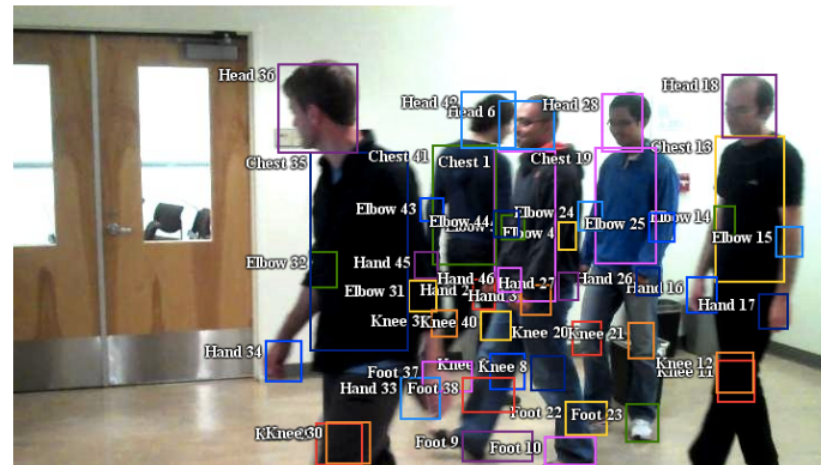
# VATIC Studies: Task Design

Subject	Scripted				Basketball				VIRAT			
	User	Fixed	Ratio	Saved	User	Fixed	Ratio	Saved	User	Fixed	Ratio	Saved
A	599	<b>463</b>	0.77	136	1,457	<b>1,323</b>	0.91	134	<b>220</b>	244	1.11	-24
B	653	<b>247</b>	0.38	406	4,555	<b>2,275</b>	0.50	2,280	<b>176</b>	<b>178</b>	1.03	-2
C	476	<b>275</b>	0.58	201	1,216	<b>830</b>	0.68	386	338	<b>215</b>	0.64	123
D	772	<b>432</b>	0.56	340	1,505	<b>1,497</b>	0.99	8	489	<b>302</b>	0.62	187
*E	605	<b>371</b>	0.61	234	<b>935</b>	1501	1.61	-566	269	<b>231</b>	0.85	38
*F	654	<b>472</b>	0.72	182	<b>1,672</b>	1,858	1.11	-186	372	<b>326</b>	0.87	46
*G	235	<b>193</b>	0.82	42	<b>591</b>	696	1.18	-105	165	<b>120</b>	0.73	45
*H	<b>312</b>	331	1.06	-19	<b>656</b>	748	1.14	-92	172	<b>164</b>	0.95	8
Mean	538	<b>348</b>	0.66	190	1,573	<b>1,341</b>	0.96	232	275	<b>223</b>	0.83	53

Fixed interval key frames significantly faster than user defined key frames!

# VATIC Studies: Task Design

Subject	Human Joints						Basketball						Preferred	Order
	One	Group		All		One	Group		All					
A	<b>1639</b>	1.00	1763	1.07	-	-	<b>476</b>	1.00	358	0.75	<b>253</b>	0.53	One	OGA
B	<b>2071</b>	1.00	2100	1.01	2440	1.17	<b>1382</b>	1.00	1618	1.17	2403	1.73	One	AOG
C	<b>1867</b>	1.00	2106	1.12	-	-	<b>681</b>	1.00	813	1.19	<b>599</b>	0.87	One	GAO
D	<b>2399</b>	1.00	-	-	-	-	<b>494</b>	1.00	<b>448</b>	0.91	960	1.94	One	AOG
Mean	<b>1994</b>	1.00	1989	1.07	2440	1.17	<b>759</b>	1.00	809	0.99	1085	1.12		



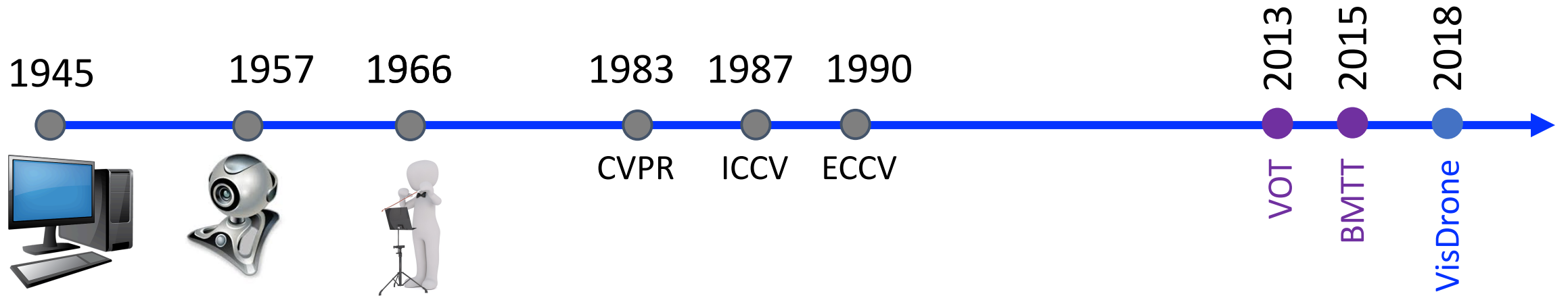
**Annotating one object at a time is more efficient and preferred!**

Carl Vondrick, Donald Patterson, and Deva Ramanan. Efficiently Scaling Up Crowdsourced Video Annotation: A Set of Best Practices for High Quality, Economical Video Labeling. IJCV 2012.

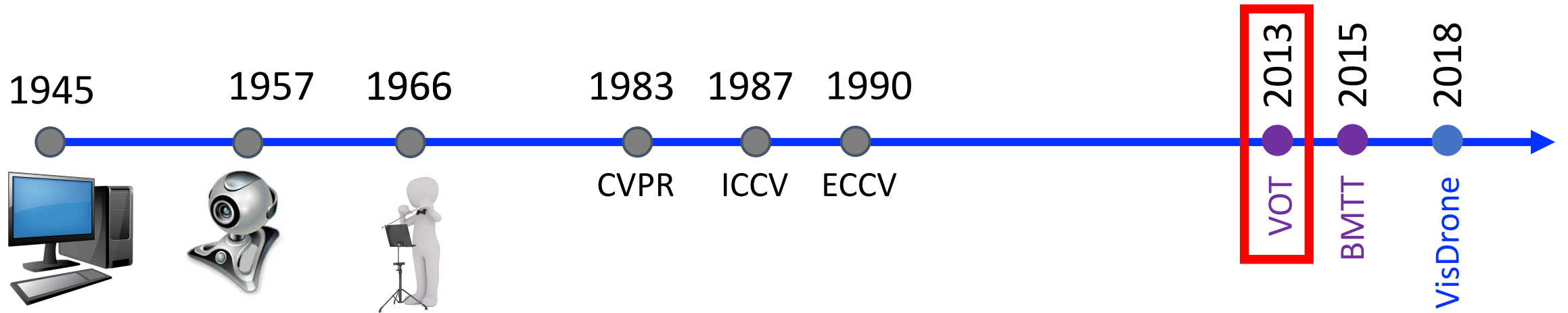
# Today's Topics

- Object tracking applications
- Evaluating object tracking methods
- Crowdsourcing object tracking strategies
- **Object tracking datasets & challenges**
- Lab: video annotation & creating figures/tables in latex

# Object Tracking Datasets



# Object Tracking Datasets

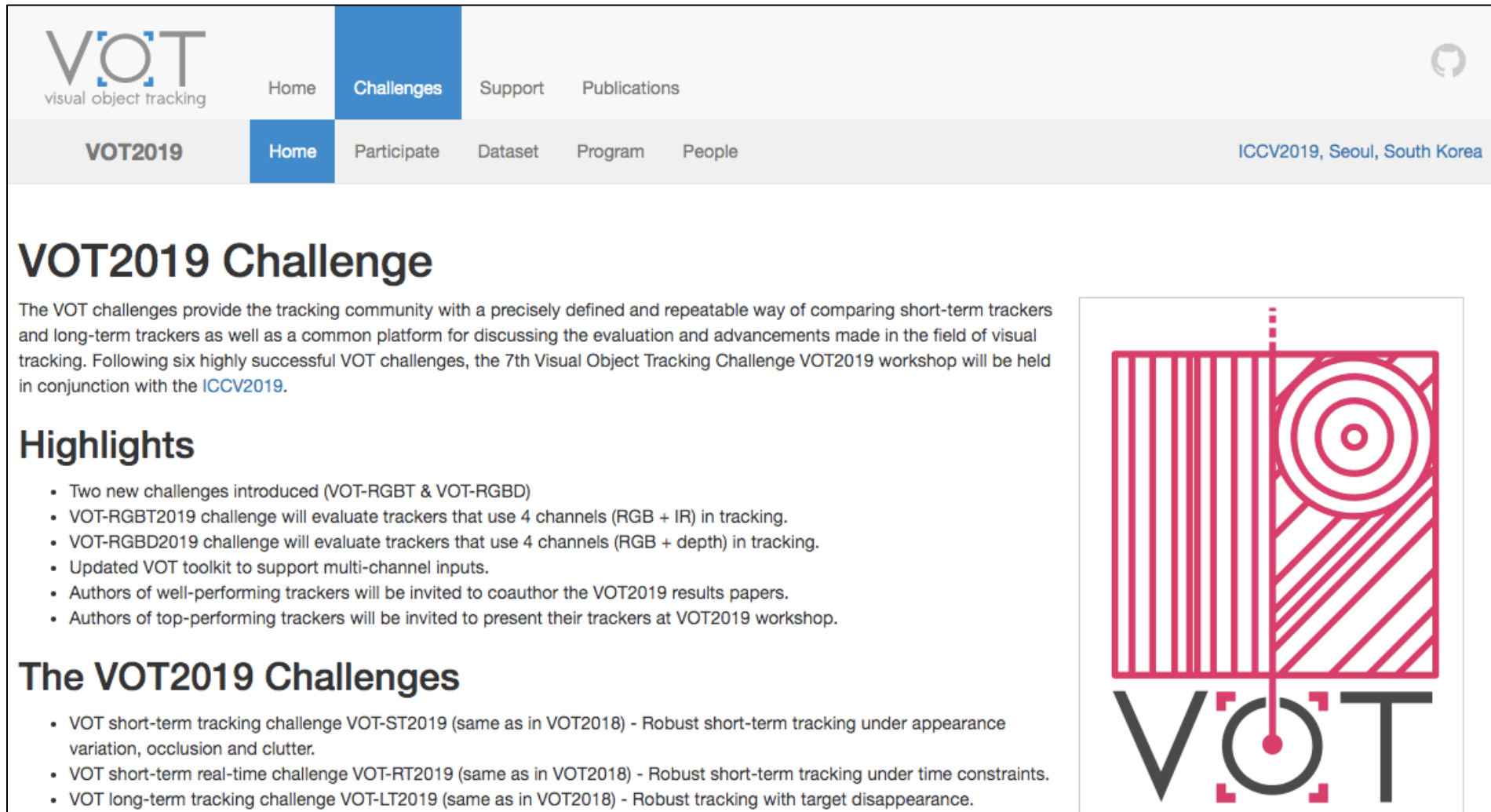




# Object Tracking Datasets: VOT

- Collected existing datasets, which resulted in inconsistent annotation methodologies across videos (e.g., different bounding box criteria)
- Authors re-annotated object tracking for videos they deemed to have unsuitable annotations

# VOT Annual Challenge (7<sup>th</sup> year now)



**VOT**  
visual object tracking

Home Challenges Support Publications

VOT2019 Home Participate Dataset Program People

ICCV2019, Seoul, South Korea

## VOT2019 Challenge


The VOT challenges provide the tracking community with a precisely defined and repeatable way of comparing short-term trackers and long-term trackers as well as a common platform for discussing the evaluation and advancements made in the field of visual tracking. Following six highly successful VOT challenges, the 7th Visual Object Tracking Challenge VOT2019 workshop will be held in conjunction with the [ICCV2019](#).

### Highlights

- Two new challenges introduced (VOT-RGBT & VOT-RGBD)
- VOT-RGBT2019 challenge will evaluate trackers that use 4 channels (RGB + IR) in tracking.
- VOT-RGBD2019 challenge will evaluate trackers that use 4 channels (RGB + depth) in tracking.
- Updated VOT toolkit to support multi-channel inputs.
- Authors of well-performing trackers will be invited to coauthor the VOT2019 results papers.
- Authors of top-performing trackers will be invited to present their trackers at VOT2019 workshop.

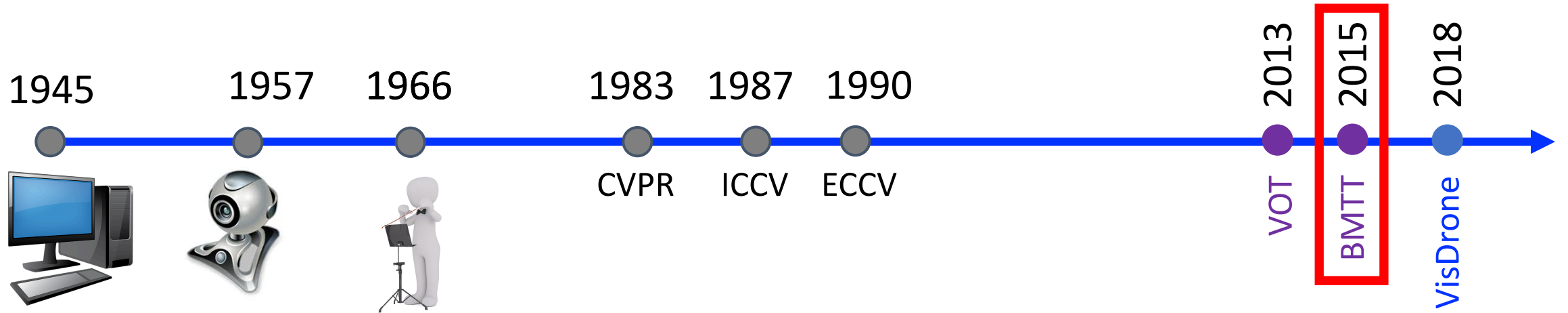
### The VOT2019 Challenges

- VOT short-term tracking challenge VOT-ST2019 (same as in VOT2018) - Robust short-term tracking under appearance variation, occlusion and clutter.
- VOT short-term real-time challenge VOT-RT2019 (same as in VOT2018) - Robust short-term tracking under time constraints.
- VOT long-term tracking challenge VOT-LT2019 (same as in VOT2018) - Robust tracking with target disappearance.



<http://www.votchallenge.net/vot2019/>

# Object Tracking Datasets



# Object Tracking Datasets: BMTT

- Authors collected 22 videos



- Unspecified how object tracking annotations were collected

# BMTT MOT Annual Challenge

## Multiple Object Tracking Benchmark

home data results vis QVA submit FAQ people login sign up

### Welcome to the MOT Challenge: The Multiple Object Tracking Benchmark!

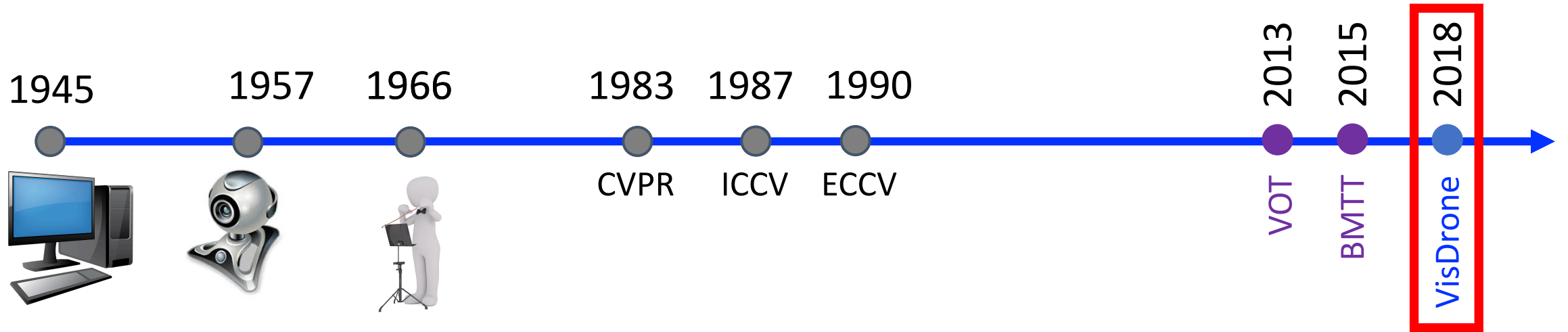
- 2D MOT 2015
- 3D MOT 2015
- MOT16
- MOT17Det
- MOT17
- DevKit



In the recent years, the computer vision community has relied on several centralized benchmarks for performance evaluation of numerous tasks including object detection, pedestrian detection, 3D reconstruction, optical flow, single-object short-term tracking, and stereo estimation. Despite potential pitfalls of such benchmarks, they have proved to be extremely helpful to advance the state-of-the-art in the respective research fields. Interestingly, there has been rather limited work on the standardization of multiple target tracking evaluation. One of the few exceptions is the well-known [PETS](#) dataset, targeted primarily at surveillance applications. Even for this widely used benchmark, a common technique for presenting tracking results to date involves using different subsets of the available data, inconsistent model training and varying evaluation scripts. With this benchmark we would like to pave the way for a unified framework towards more meaningful quantification of multi-target tracking.

<https://motchallenge.net/>

# Object Tracking Datasets



# Object Tracking Datasets: VisDrone

- Authors collected 263 video clips (179,264 frames) from drones in Asia



- Annotations created for over 2.5 million object instances, however it is unspecified how these annotations were collected



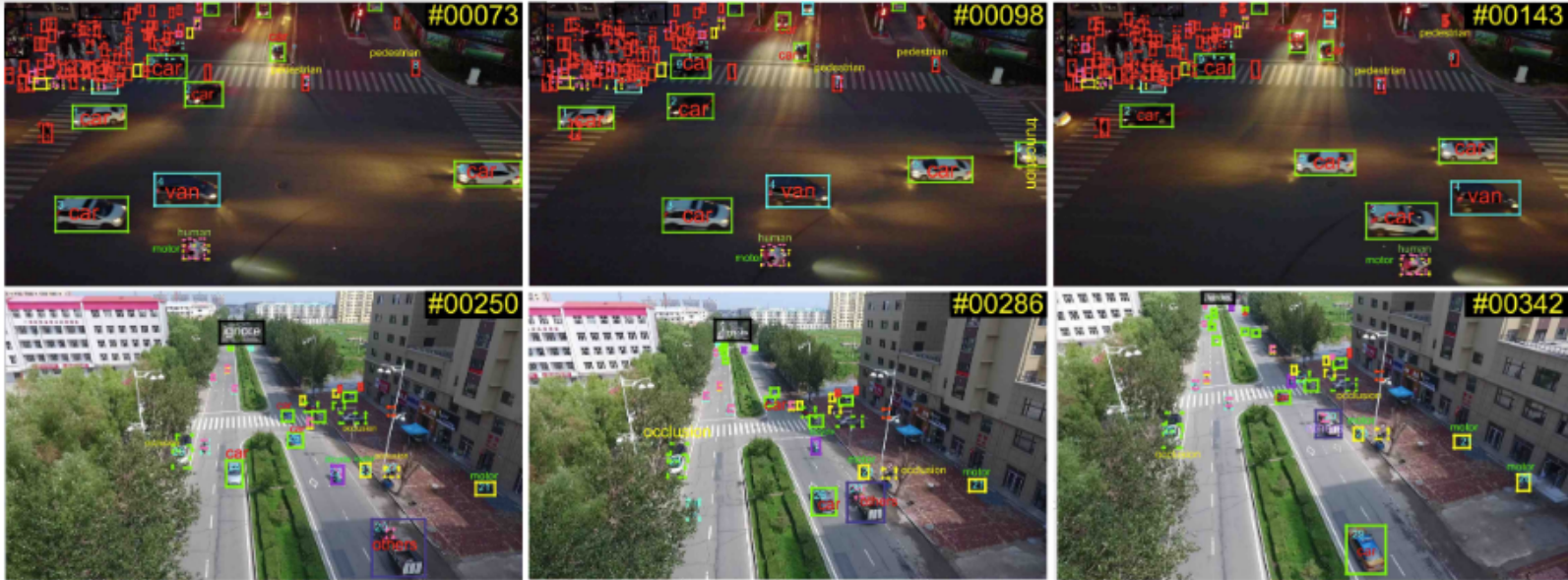
# VisDrone Challenge

Home Challenge Evaluate Download Submit FAQ ICCV2019 People Sign in Sign up

Object Detection in Images  
Object Detection in Videos  
**Single-Object Tracking**  
Multi-Object Tracking

## Multi-Object Tracking

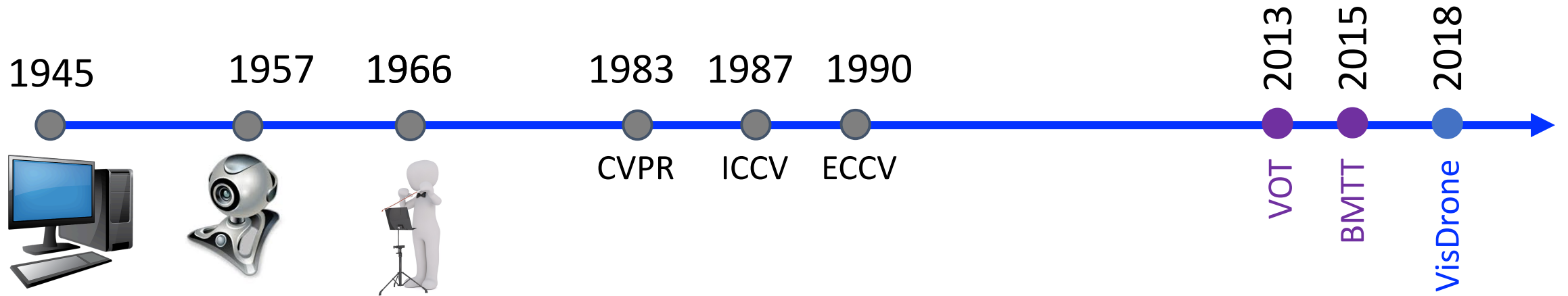
...r results here!! Note that the evaluation server on the test-dev set will be open for



<http://www.aiskyeye.com/views/index>



# Object Tracking Datasets



# Discussion

- Design a crowdsourcing system to collect **high quality** object tracking annotations. Specify how your system design will handle difficult scenarios such as when an object is:
  - Partially visible
  - Occluded
  - Is reflected in reflective surfaces such as mirrors or windows
- What will be the total crowdsourcing task cost if we use your system to annotate 1,000 1-minute videos where you need to track 5 humans/video (assume 30 frames/second)?

# Today's Topics

- Object tracking applications
- Evaluating object tracking methods
- Crowdsourcing object tracking strategies
- Object tracking datasets & challenges
- **Lab: video annotation & creating figures/tables in latex**