NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

NIST INFORMATION TECHNOLOGY LABORATORY

TRECVID22 Activity Detection in Extended Video Task Summary

Presenter: Jon Fiscus*

Yooyoung Lee *, Andrew Delgado@, Afzal Godil *,
Baptiste Chocot +, Lukas Diduch #,
Eliot Godard +, Jim Golden *

* NIST Staff
+ NIST Guest Research
# Dakota Consulting
@ Former NIST Staff

Day 3
December 8
9:20 a.m. – 10:40 a.m.(ET)

Discussions on Slack

Disclaimer

Certain commercial equipment, instruments, software, or materials are identified in this paper to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor necessarily the best available for the purpose.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, NIST, or the U.S. Government.

Outline

- Self Reported Leaderboard (SRL) Evaluation Overview
- Evaluation Tasks and Measures
- Training and Testing Datasets
- Results and Analyses
- Next Steps

ActEV Self Reported Leaderboard (SRL) Evaluation Overview

ActEV - Activity Detection in Extended Videos

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

What is ActEV?

- A series of challenges designed to promote
  - Robust detection of known (pre-trained) and surprise (ad-hoc) activities in
    - Known and unknown facilities
    - Multi-camera environment
  - Temporal and spatio-temporal localization of the activity for reasoning

- Brief History
  - 2018-2021 – Sequestered Data Evaluations for the IARPA DIVA Program
  - 2022-Beyond – Continued research leveraging Gov't collected sequestered data

ActEV SRL Evaluation Structure

- Two retrospective analysis evaluation tasks
  - Primary – Activity and Object Detection (AOD)
    - Detection and Spatial-Temporal Localization
  - Secondary – Activity Detection (AD)
    - Detection and Temporal Localization

- Evaluation Type
  - Self-reported (i.e., Take-Home) evaluation
    - Participants download the test dataset, run their systems on their machines, and submit the system outputs to NIST

ActEV'22 Tasks and Measures

Evaluation Tasks

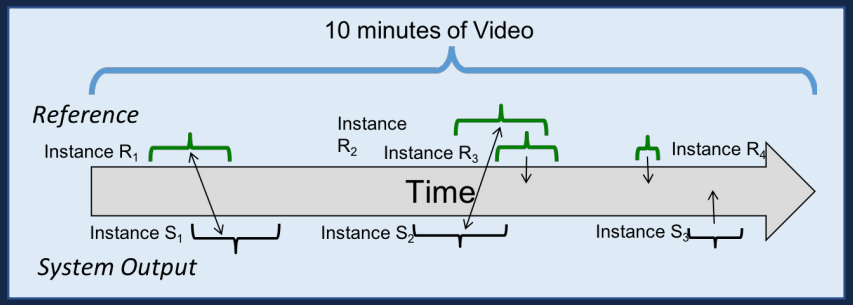
Activity Detection (AD)
Detection and
temporal localization

Primary

Activity and Object Detection (AOD)
Detection and
spatio-temporal localization

NIST IRB #ITL-17-0037

Detection Performance Measure



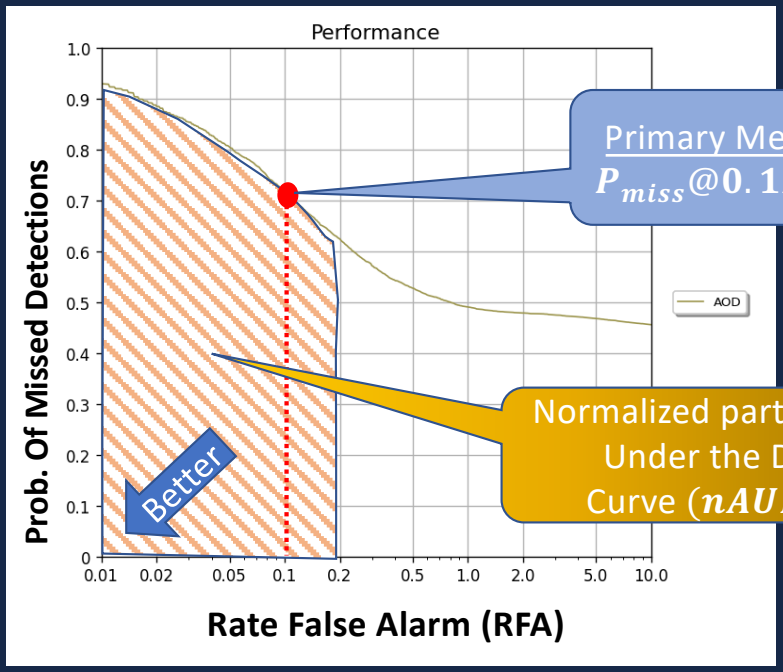
Detection Performance

$$P_{miss}(\tau) = \frac{\#_{Missed}(\tau)}{\#_{TrueInstance}}$$

$$\text{RFA}(\tau) = \frac{\#_{FalseAlarm}(\tau)}{VideoDurInMinutes}$$

$$\tau = is\ a\ presence\ confidence\ treshhold$$

DET (Detection Error Tradeoff)



Averaged over activities and measured at specific RFA values

Please see details at https://actev.nist.gov/sdl

Spatio-Temporal Localization Performance
(a Secondary Measure on Correctly Detected Activity Instances)

N_MODE (Normalized Multiple Object Detection Error)

$$N_{MODE}(\tau) = \sum_{t=1}^{N_{frames}} \frac{\left(MD_t(\tau)+FA_t(\tau)\right)}{N_{frames}}$$

Ratio of errors to total frames (N_{frames})

- $MD_t(\tau)$ = Missed Frame Localization
  - The number of frames where the system did not localize the box at presence confidence τ
- $FA_t(\tau)$ = False Alarm Frame Localization
  - The number of frames where the system added a incorrect box at presence confidence τ

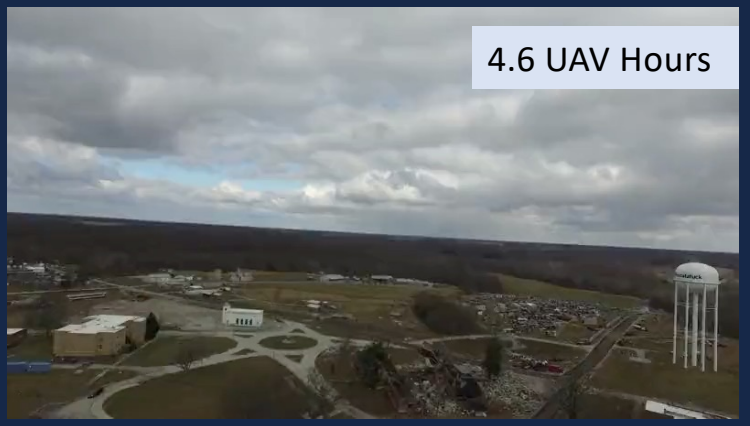
TRECVID ‘22 ActEV Dataset

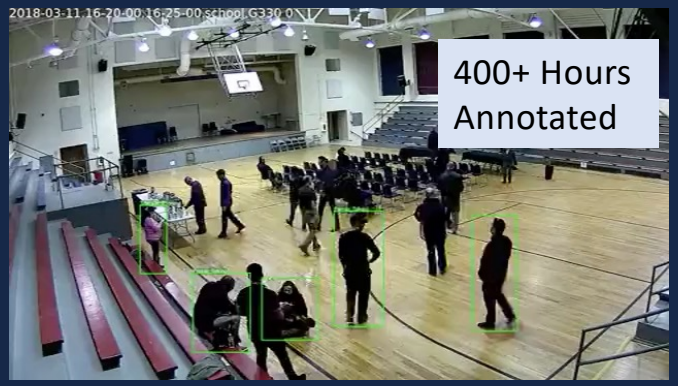
Approved by Institutional Review Board (IRB)

#ITL-17-0037

https://mevadata.org/

Training and Evaluation Data from the Multiview Extended Video with Activities (MEVA) Dataset





NIST IRB #ITL-17-0037

20 Activities for SRL ’22

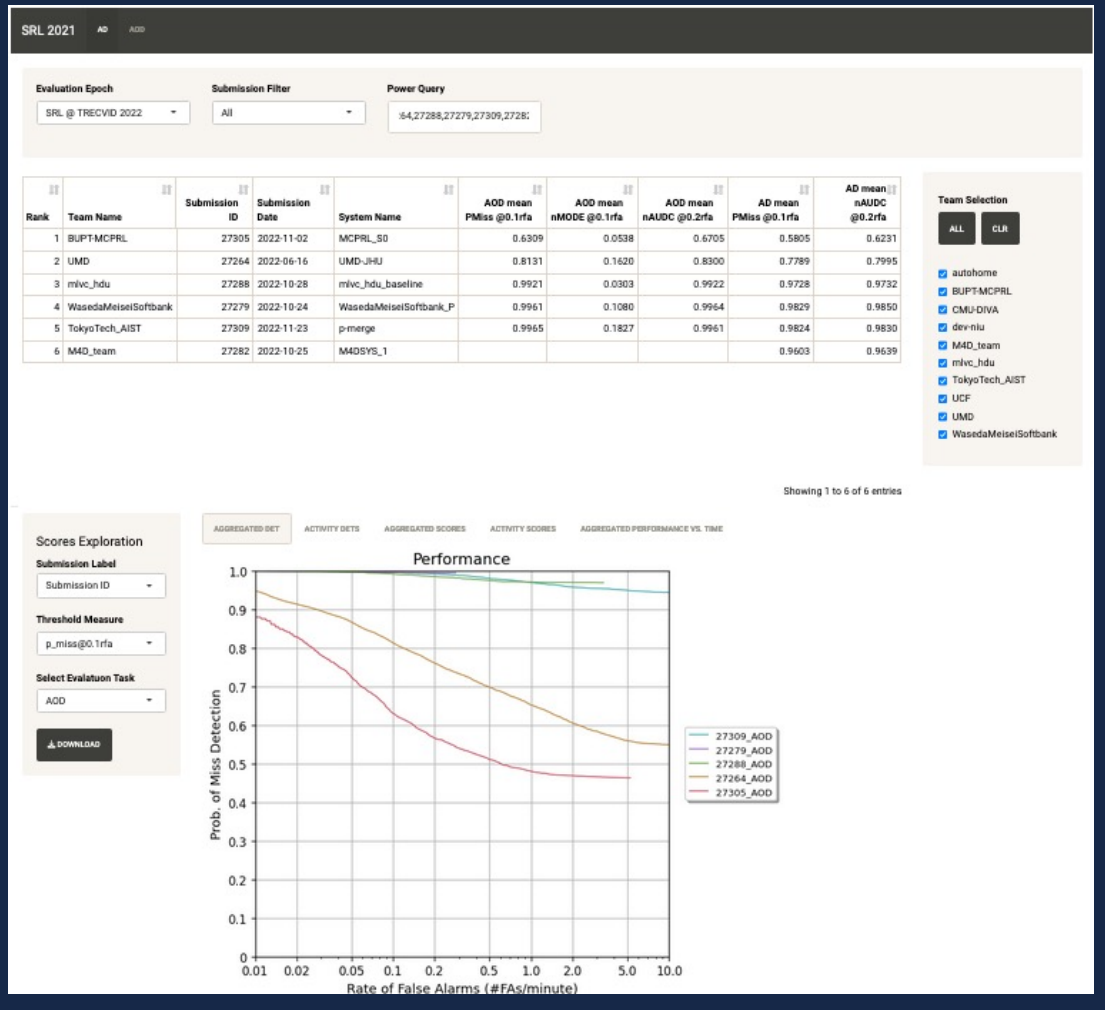
(Subset of the 37 Sequestered Data Leaderboard “Known” Activities where systems trained for the activities)

person_closes_vehicle_door
person_enters_scene_through_structure
person_enters_vehicle
person_exits_scene_through_structure
person_exits_vehicle
person_interacts_with_laptop
person_opens_facility_door
person_opens_vehicle_door
person_picks_up_object
person_puts_down_object
person_reads_document
person_sits_down
person_stands_up
person_talks_to_person
person_texts_on_phone
person_transfers_object
vehicle_starts
vehicle_stops
vehicle_turns_left
vehicle_turns_right

Results and Analyses

ActEV SRL
Results and Analyses

https://actev.nist.gov/SRL#tab_leaderboard



NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

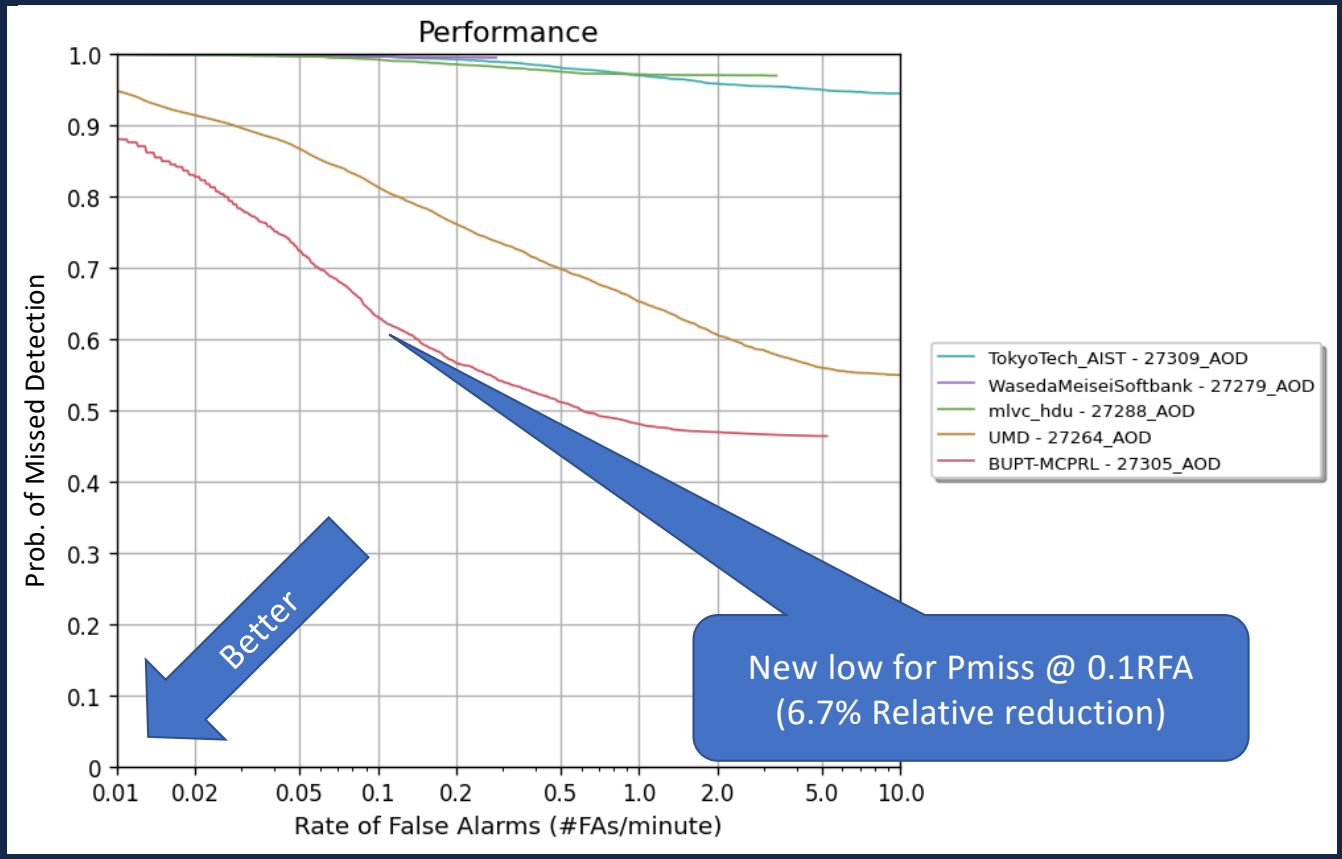
ActEV@TRECVID '22 SRL Participants Ranking

38 submissions as of 11/02/2022 4:00 PM EST (+3 late)

6 teams (top performing system result per team ordered by Pmiss@0.1RFA)

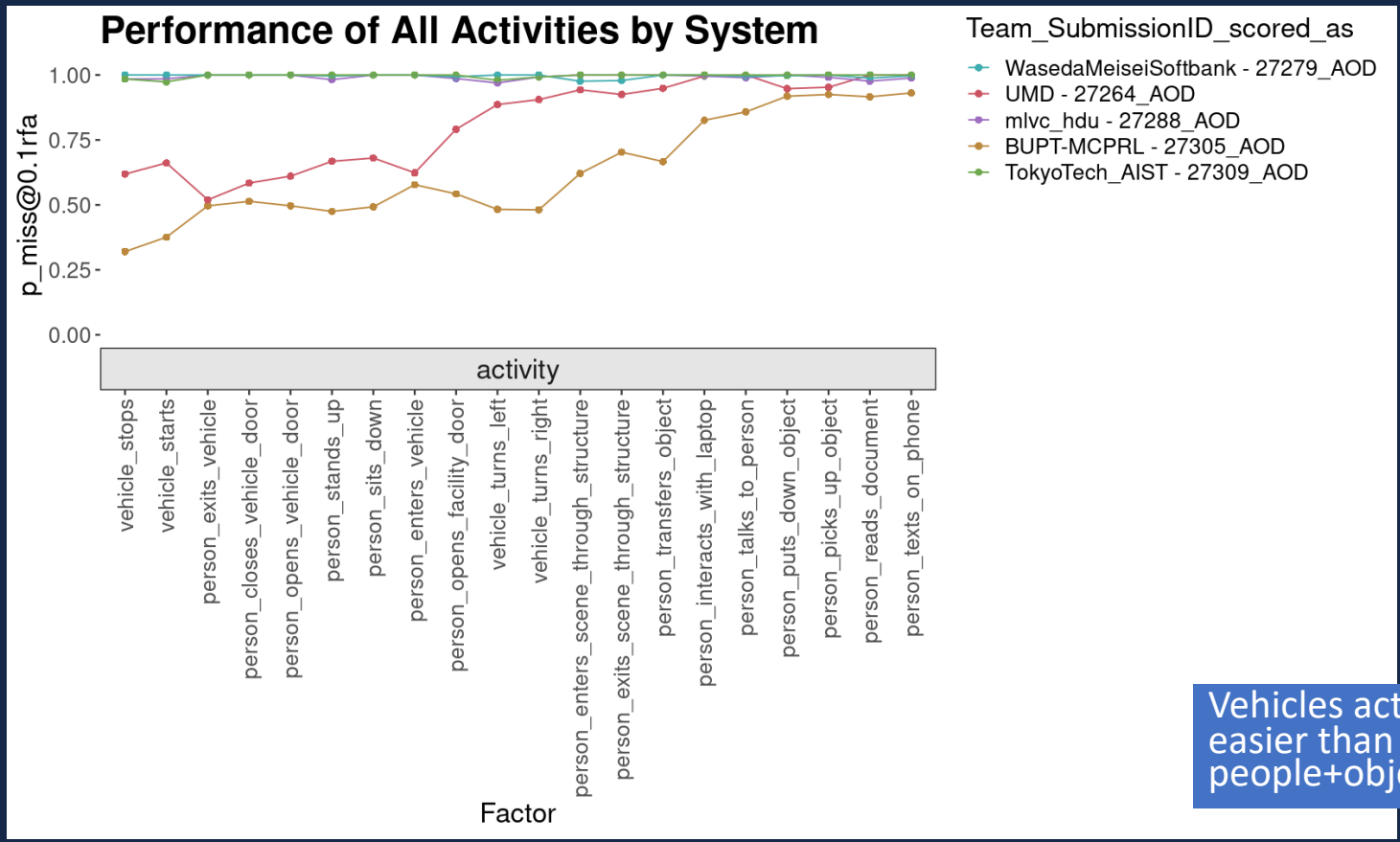
Team	Organization	Primary Task: Activity and Object Detection (AOD)			Secondary Task: Activity Detection (AD)	
		Pmiss @0.1RFA	nMODE @0.1RFA	nAUDC @0.2RFA	Pmiss @0.1RFA	nAUDC @0.2RFA
BUPT-MCPRL	Beijing University of Posts and Telecommunications, China	0.6309	0.0538	0.6705	0.5805	0.6231
UMD	University of Maryland, USA	0.8131	0.1620	0.8300	0.7789	0.7995
mlvc_hdu	Hangzhou Dianzi University	0.9921	0.0303	0.9922	0.9728	0.9732
Waseda Meisei Softbank	Waseda University, Meisei University, SoftBank Corporation	0.9961	0.1080	0.9964	0.9829	0.9850
TokyoTech AIST (late)	Tokyo Institute of Technology	0.9965	0.1827	0.9961	0.9824	0.9830
M4D_team	Centre for Research and Technology Hellas				0.9603	0.9639

ActEV SRL
Activity and Object Detection (AOD) DET Curve



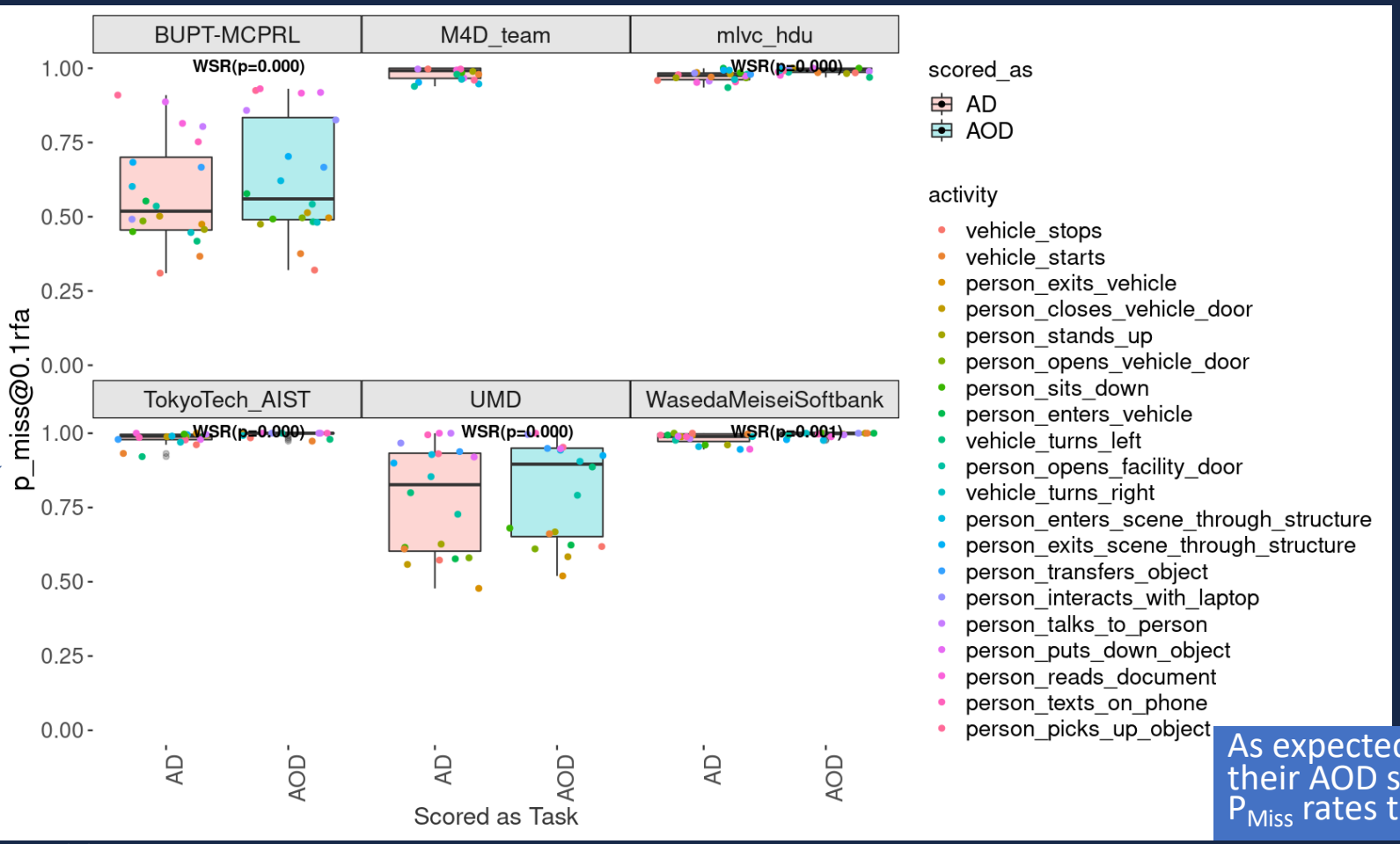
ActEV SRL
AOD Activity Specific Performance

Better



Vehicles activities remain easier than people and people+object activities

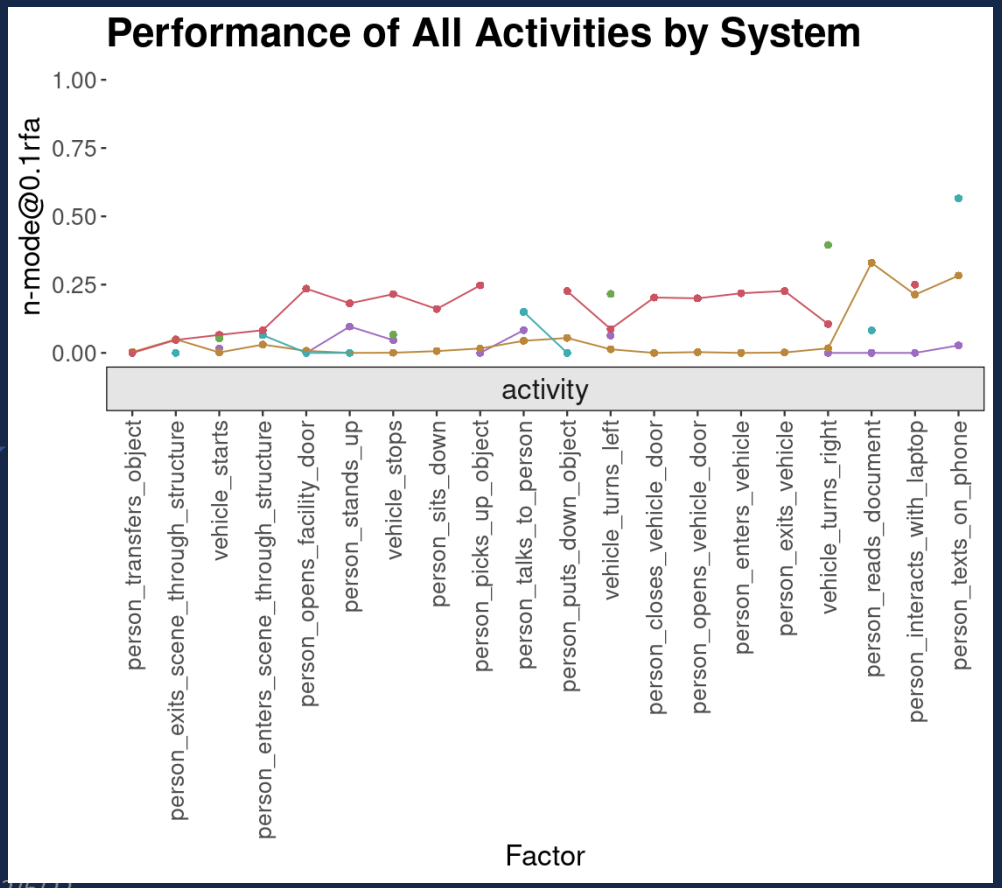
AD vs. AOD Detection Performance



Better

As expected for every team, their AOD system has higher P_{Miss} rates than AD

Localization Performance for Correct AOD Instances



Better

Team_SubmissionID_scored_as
- WasedaMeiseiSoftbank - 27279_AOD
- UMD - 27264_AOD
- mlvc_hdu - 27288_AOD
- BUPT-MCPRL - 27305_AOD
- TokyoTech_AIST - 27309_AOD

- Localization performance varies across teams
- Missing points indicate no correct detections
- BUPT-MCPRL localizes well for most activities

Summary

- This was the first use of MEVA data in TRECVID
  - The test dataset was also used for WACV'22 ActEV SRL evaluation and the ActivityNet ActEV Challenge.
- First ActEV test for MLVC_HDU, TokyoTech, and WadsedaMeiselSoftbank
- Detection and Localization (AOD) remains a more difficult task

ActEV Next Steps

- The evaluation set was used for three evaluation/workshops
  - What should be next for TRECVID?
  - What is the present data pain-point/deficiency
- Future activity detection technology challenges to investigate
  - Unknown camera views – MEVA video from new cameras
  - Infrared video – video from IR cameras
  - Unknown facility – video from new collection site
  - Surprise activities – zero shot and minimal video training
- Alternative technology challenge opportunities
  - Person tracking
  - Person re-identification

NIST | NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST | INFORMATION TECHNOLOGY LABORATORY

Thank You

ActEV Agenda

Time	Title	Presenters
8:00 - 8:20am EST	**Activities in Extended Videos - Task Overview**	Jonathan Fiscus, NIST
8:20 - 8:40am EST	**An Effective Framework for Activity Detection in Untrimmed Security Videos**	Hangyue Zhao, Beijing University of Posts and Telecommunications, China
8:40 - 9:00am EST	**Human and vehicle activity detection and recognition from untrimmed videos**	Despoina Touska, Centre for Research & Technology Hellas, Information Technology Institute, CERTH-ITI
9:00 - 9:20am EST	**A proposal-based solution to spatio-temporal action detection in untrimmed videos**	Ketul Shah , Johns Hopkins University
9:20 - 9:40am EST	**Waseda_Meisei_SoftBank at TRECVID 2022: Activities in extended videos**	Hideaki Okamoto, SoftBank Corp.
9:40 - 9:50am EST	**Poster: Dynamic Interactive Aggregation Network for TRECVID'22 ActEV Task**	Xingchao Ye, Hangzhou Dianzi University, China
9:50 - 10:10am EST	**Break**	
10:10 - 10:30am EST	**Discussion**	