

# TRECVID-2015

## Concept Localization : Overview

George Awad

NIST  
Dakota Consulting, Inc

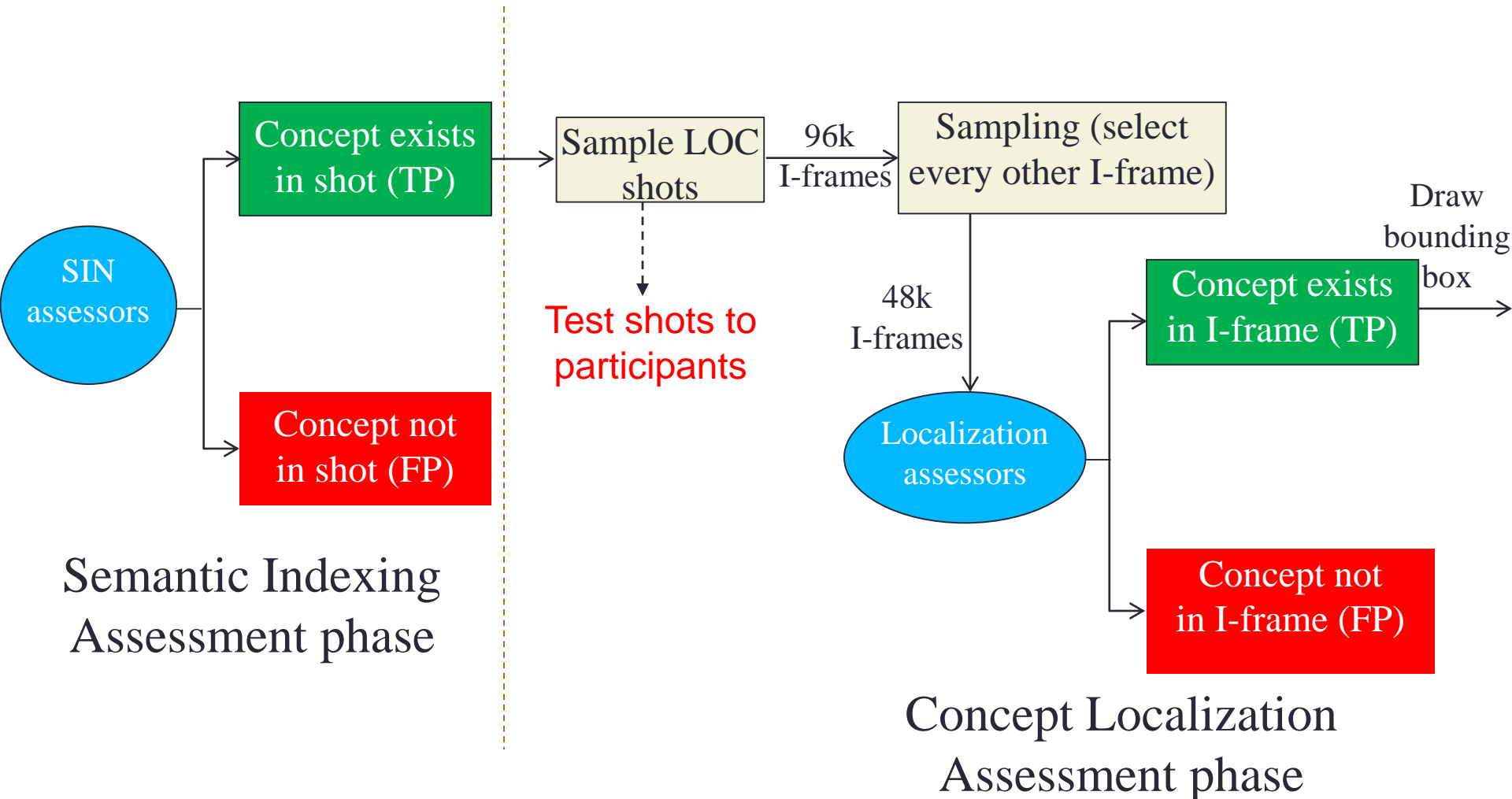
- **Goal**
  - Make concept detection more precise in **time** and **space** than current shot-level evaluation.
  - Encourage more reusable concept detectors design that is independent from the context.
- **Task**
  - This year the task is independent from SIN.
  - For each of the 10 test concepts, NIST provided set of TP shots (up to max 300).
  - For each I-Frame within the shot that contains the target, return the x,y coordinates of the (UL,LR) vertices of a bounding rectangle containing all of the target concept and as little more as possible.
- Systems were allowed to submit more than 1 bounding box per I-frame but only the one with maximum fscore were scored.

# 10 Evaluated Concepts

- Airplane
- Anchorperson\*
- Boat\_Ship
- Bridges
- Bus
- Computers\*
- Motorcycle
- Telephones
- Flags
- Quadraped

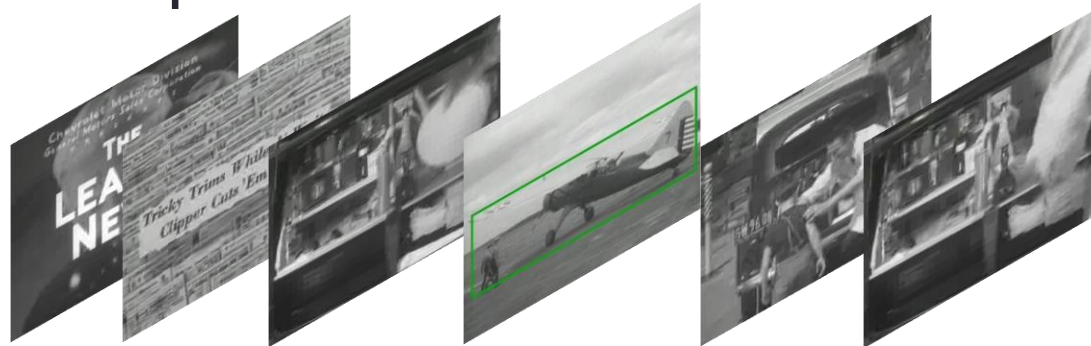
\* New concepts

# NIST Evaluation framework



# Evaluation metrics

- **Temporal localization:** precision, recall and fscore based on the judged I-frames.
- **Spatial localization:** precision, recall and fscore based on the located pixels representing the concept.
- An average of precision, recall and fscore for temporal and spatial localization across all I-frames for each concept and for each run.

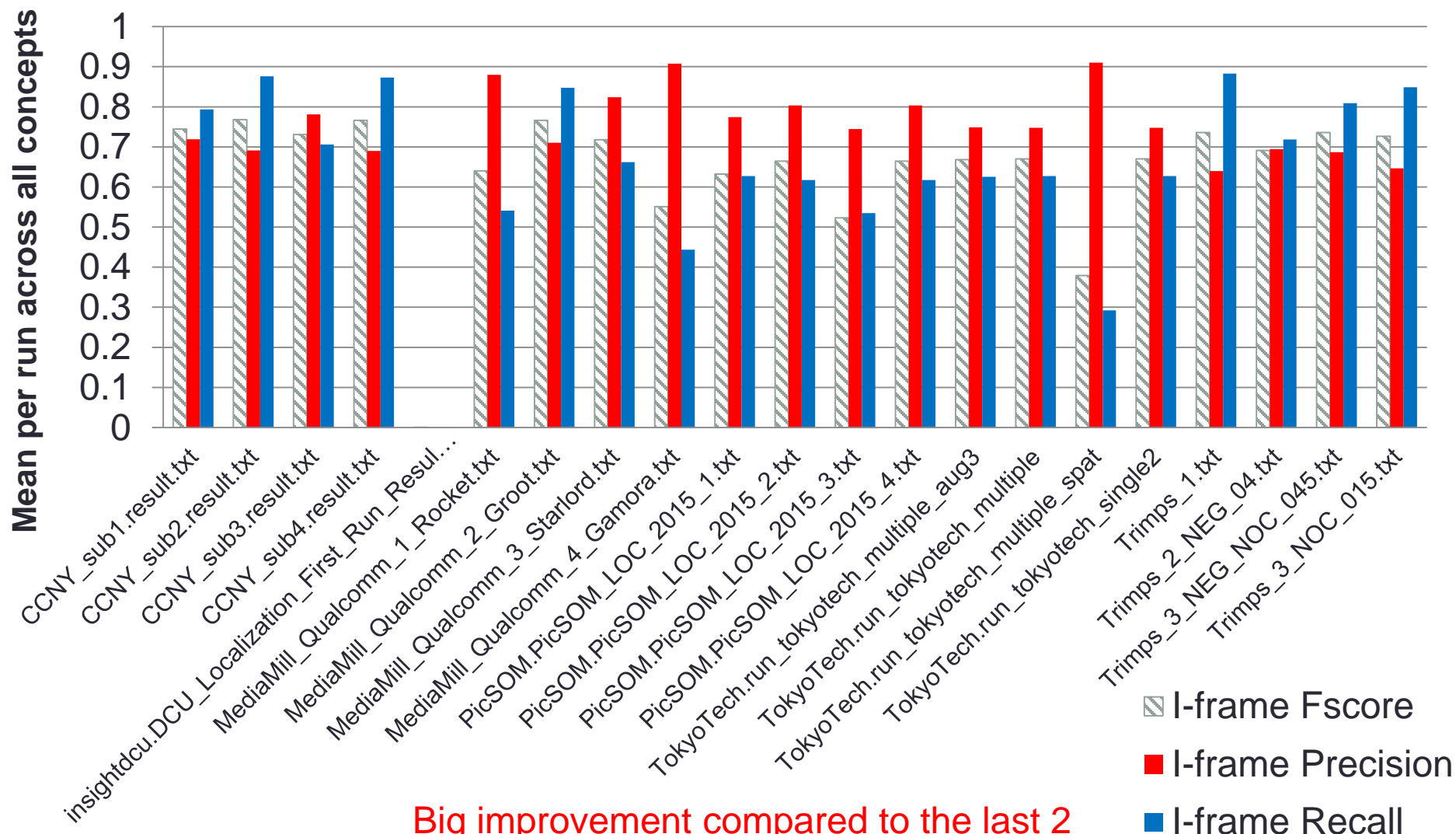


# Participants (Finishers 6 out of 18)

- 6 teams submitted 21 runs
  - MediaMill U. Of Amsterdam; Qualcomm
  - CCNY City College of New York; CUNY
  - TokyoTech Tokyo Institute of Technology
  - Trimps Third Research Institute of the Ministry of Public Security, China.
  - Insightdcu Dublin City U.; U. Polytechnica Barcelona
  - PicSom Aalto University University of Helsinki

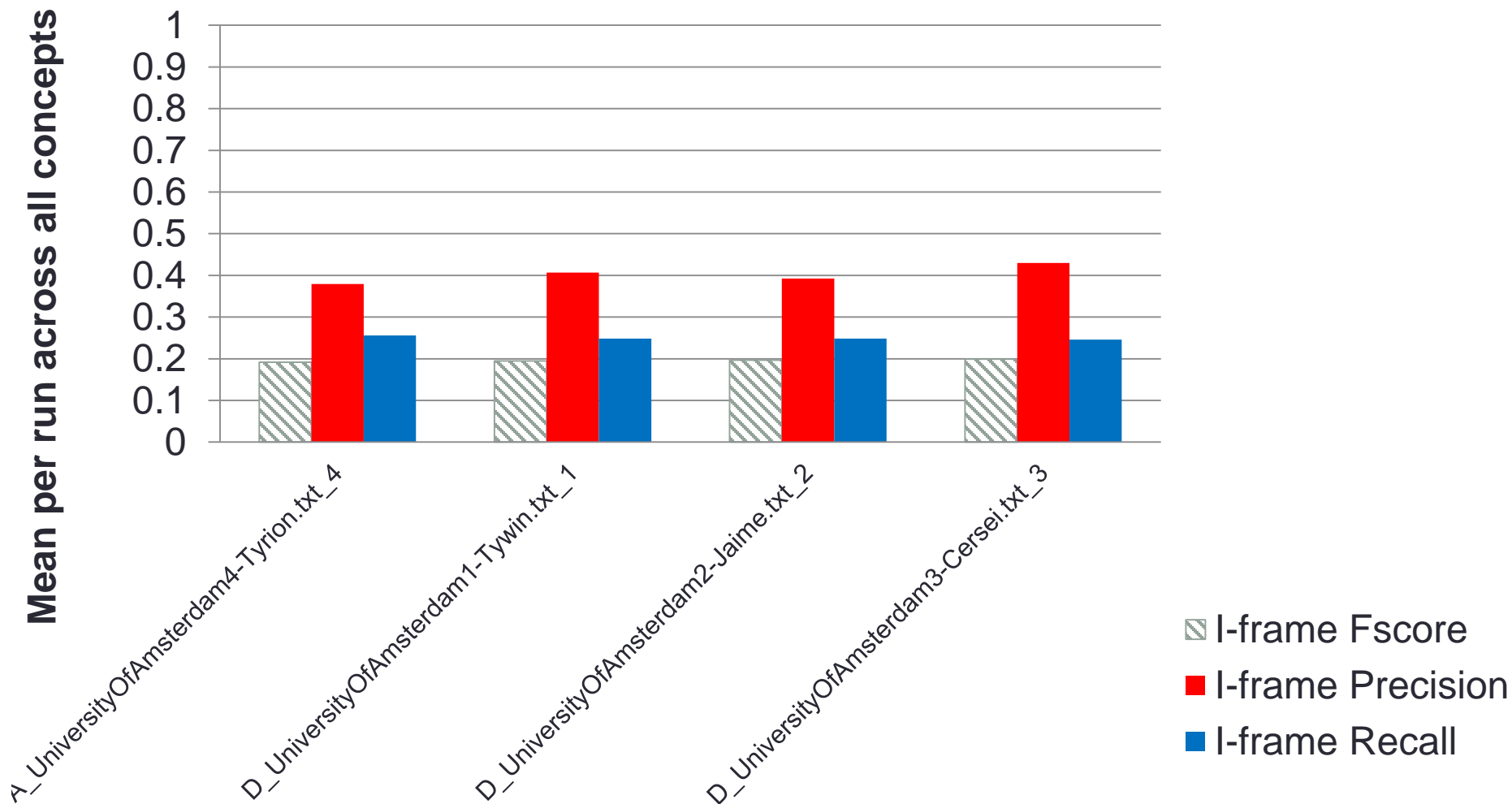
Making the task independent from SIN encouraged more participants  
(2014 : 1 team finished!)

# Temporal localization results by run



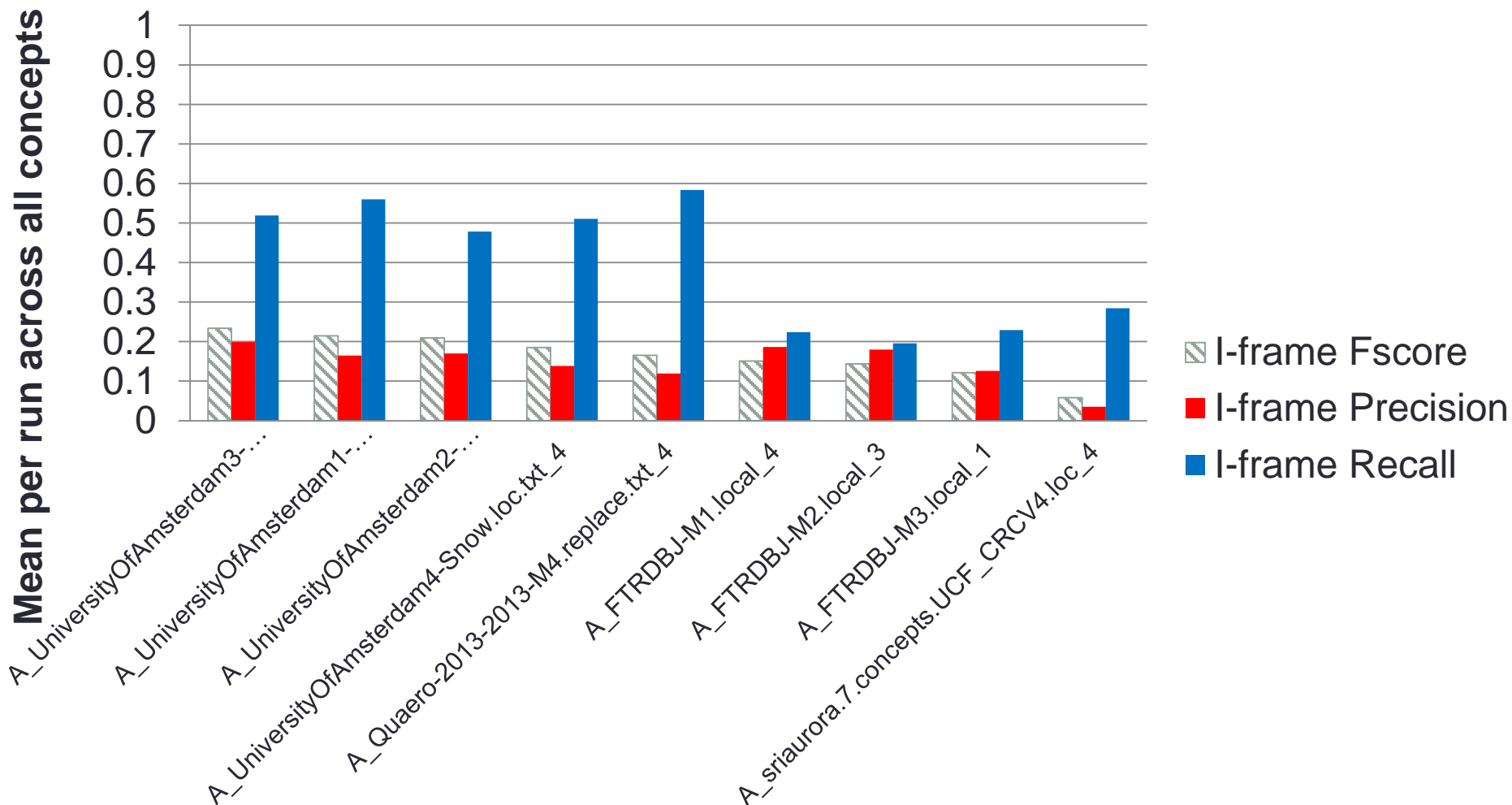
Big improvement compared to the last 2 years! probably due to focusing only on TP SIN shots

# Temporal localization results by run 2014

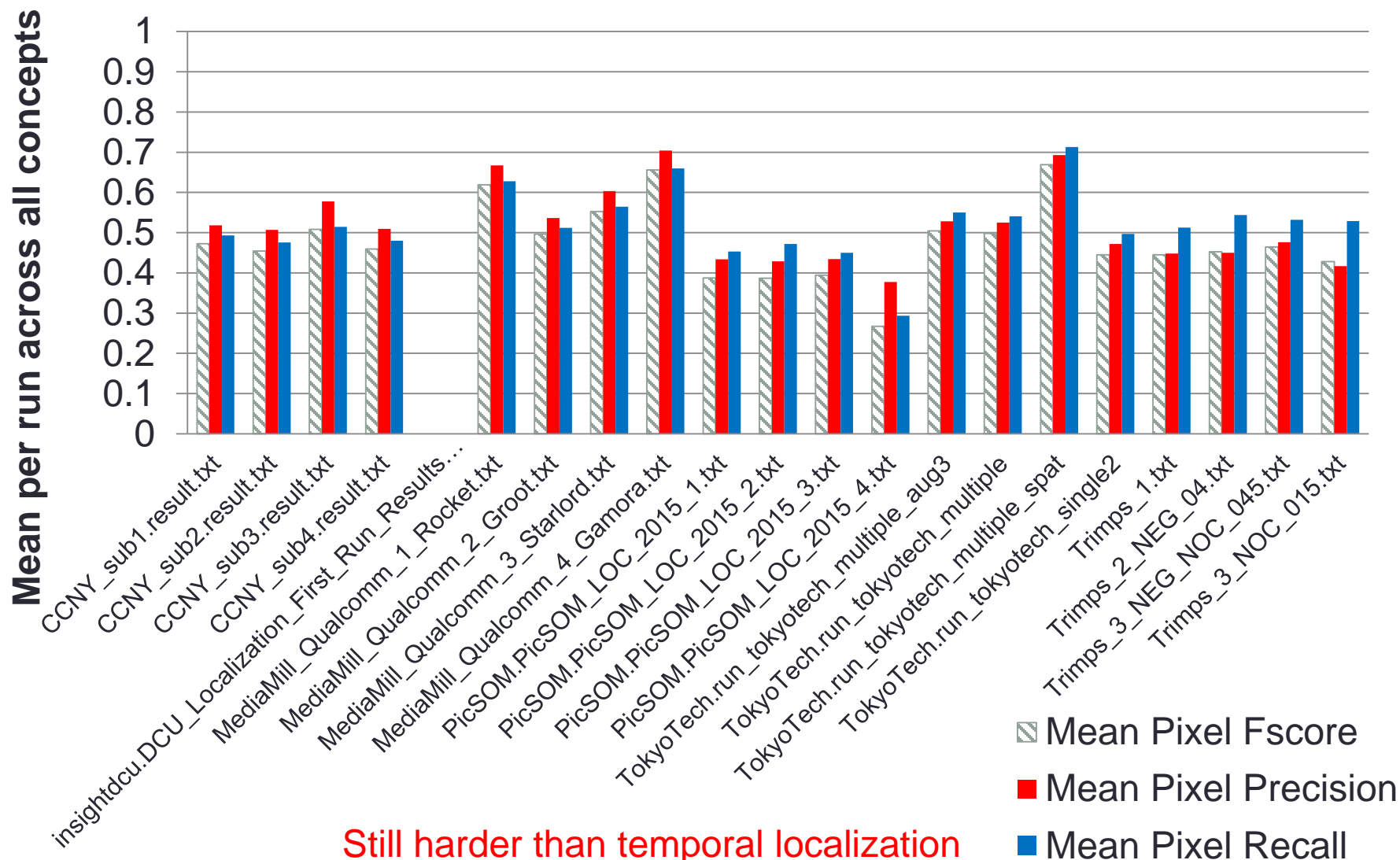




# Temporal localization results by run 2013



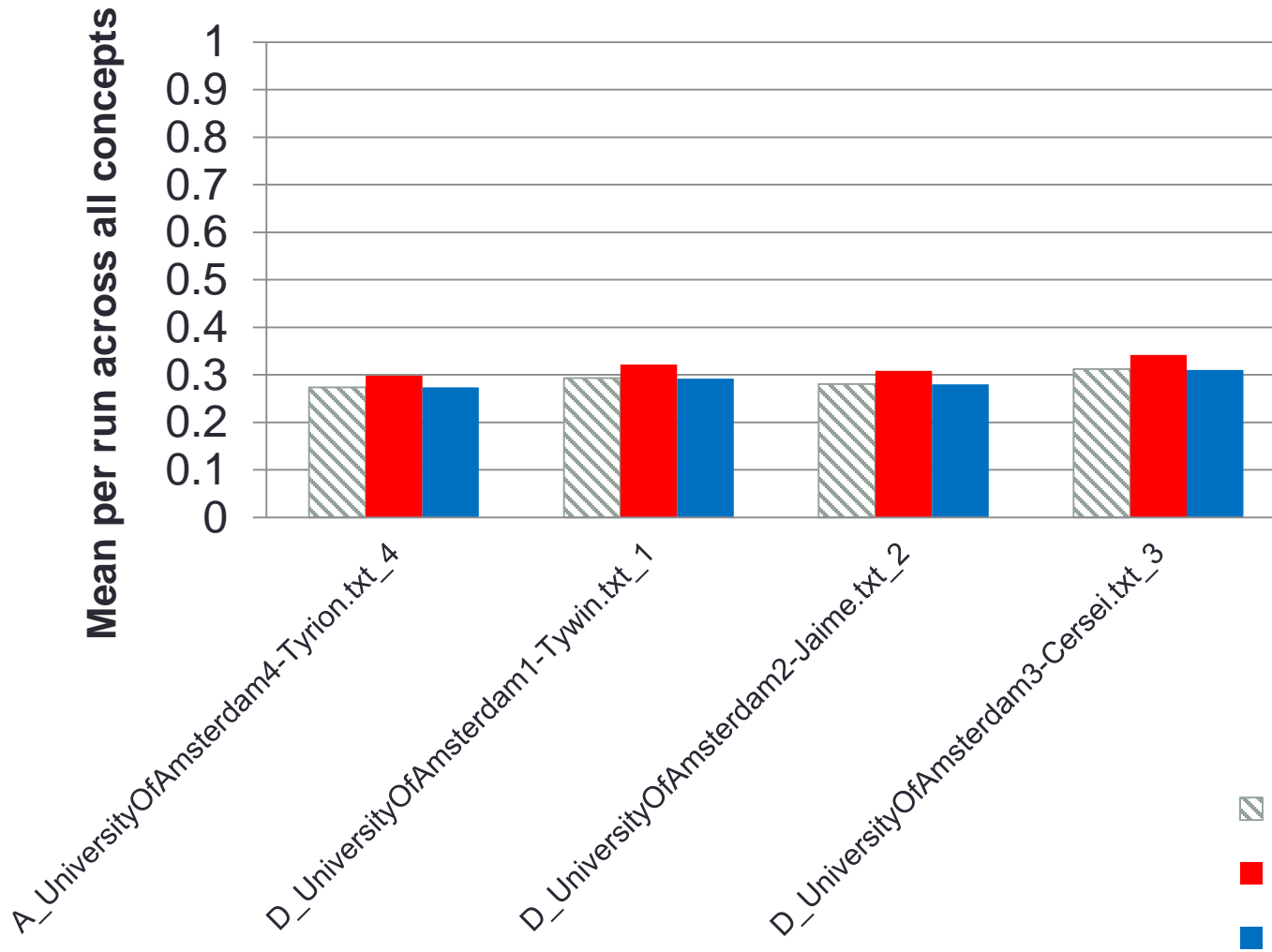
# Spatial Localization results by run



Still harder than temporal localization  
But systems improved compared to  
the last 2 years

# Spatial Localization results by run 2014

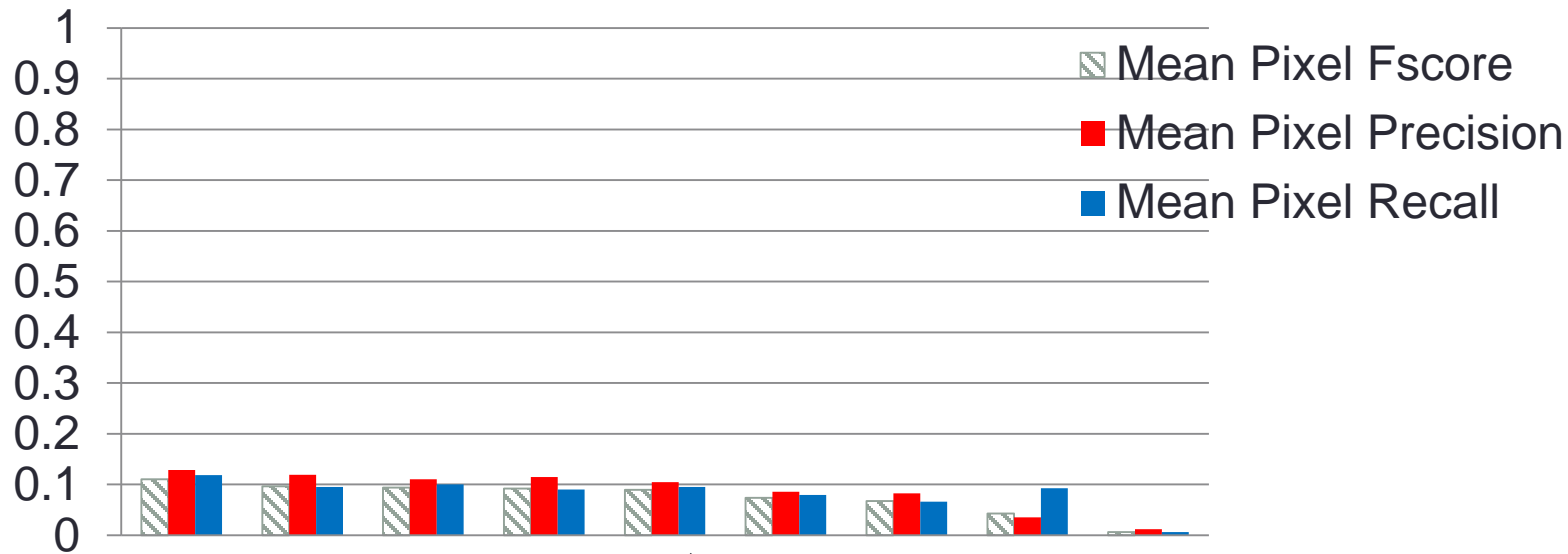
Spatial localization seems to be better than temporal (contrary to 2013 results).  
Hard to conclude as all runs come from 1 team



- ▨ Mean Pixel Fscore
- Mean Pixel Precision
- Mean Pixel Recall

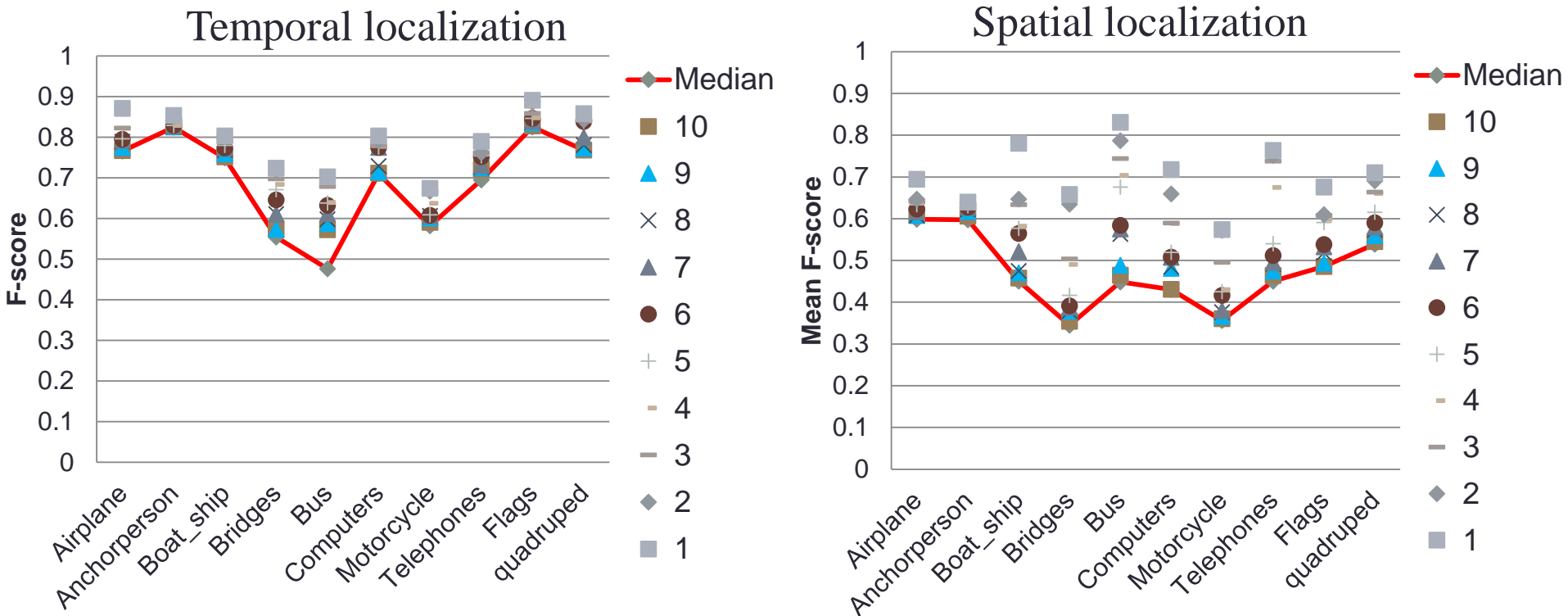
# Spatial Localization results by run 2013

Mean per run across all concepts



Finding the best bounding box is much harder than finding just the I-frame.

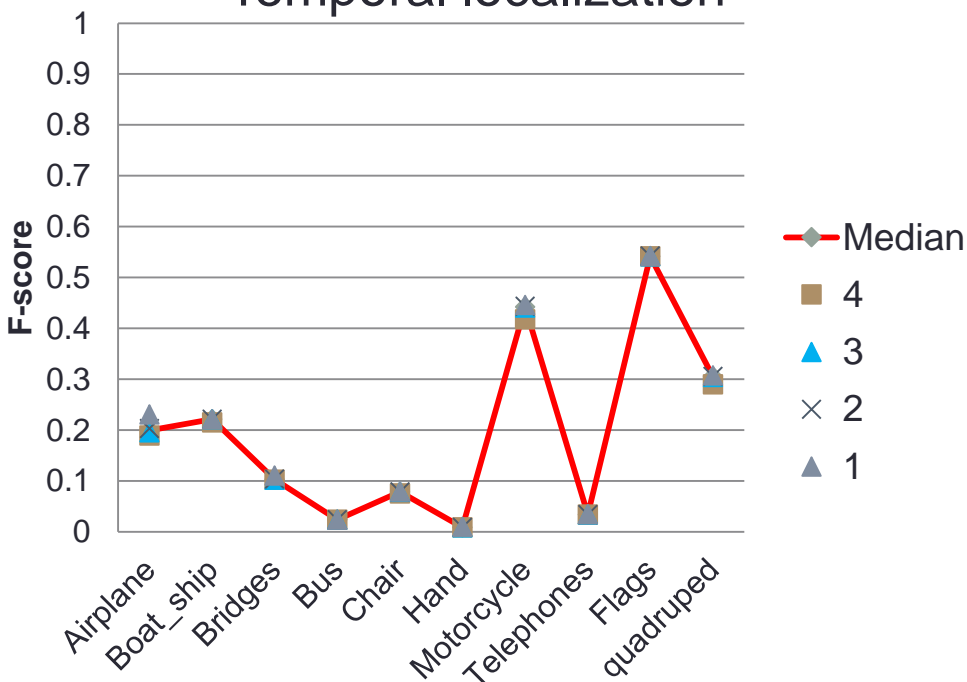
# Results per concept top 10 runs



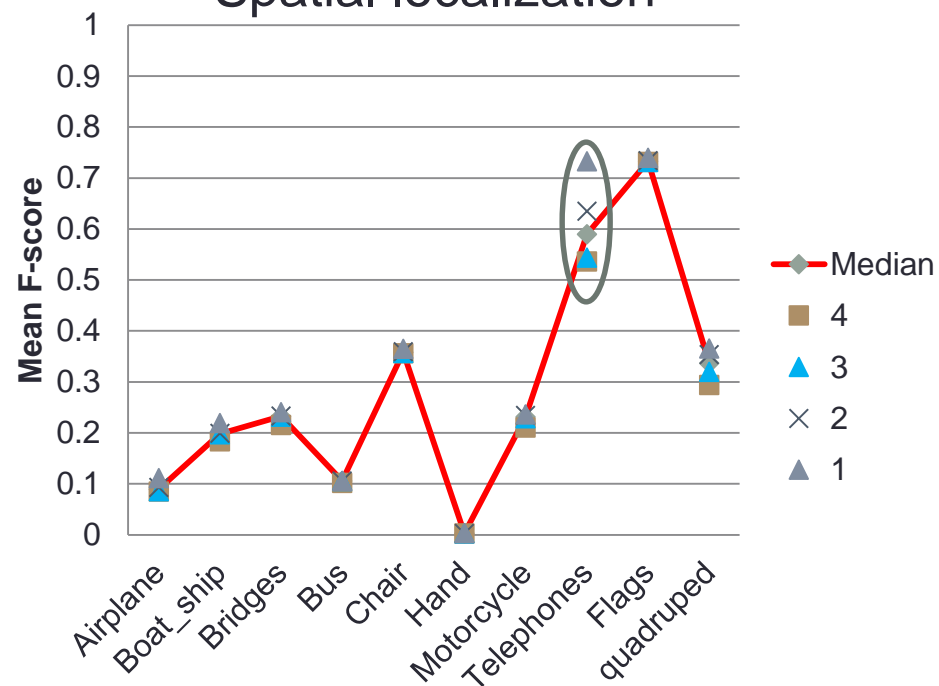
Temporal localization is easier than spatial localization  
 More variation in performance in spatial than temporal  
 Both measures are better compared to 2013 & 2014

# Results per concept 2014

## Temporal localization



## Spatial localization

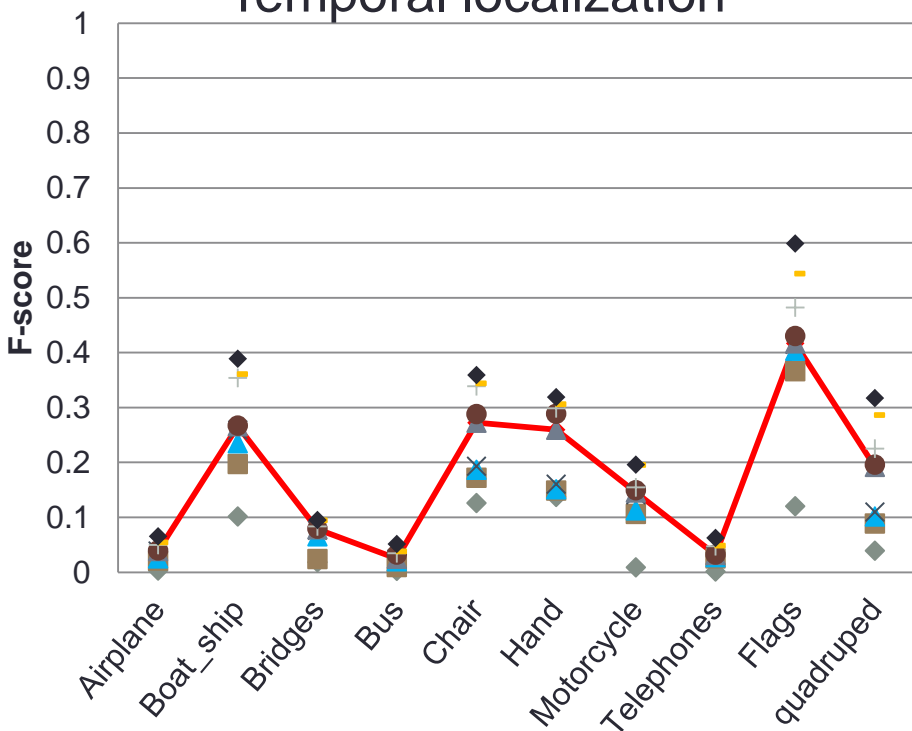


Most concepts are better in spatial localization compared to temporal. However, 1 team runs are not enough to conclude!

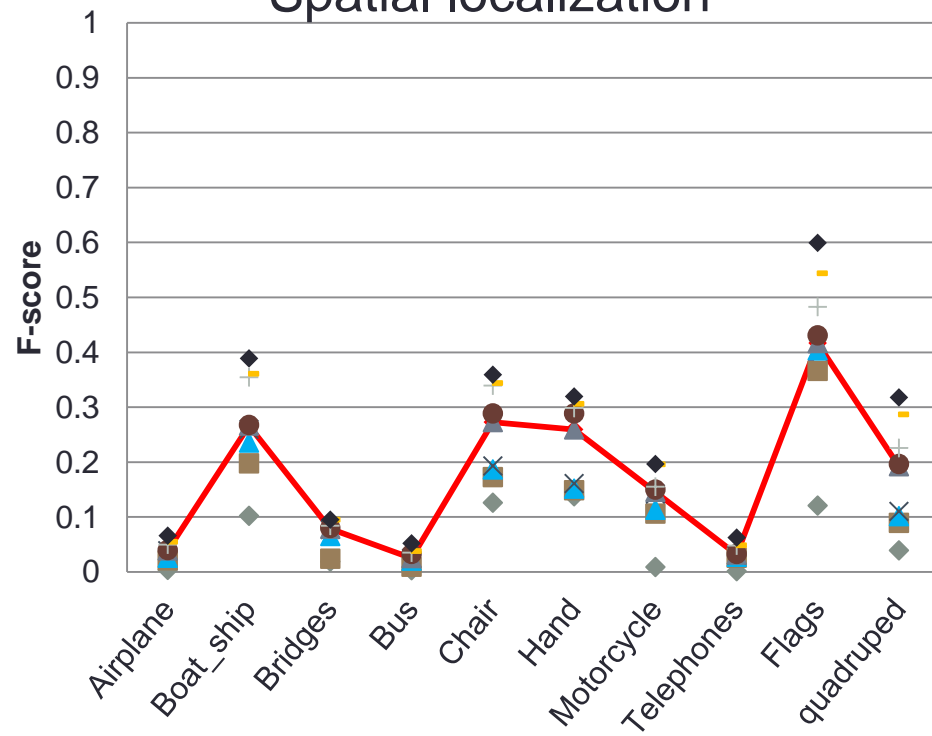
# Results per concept

## 2013

### Temporal localization

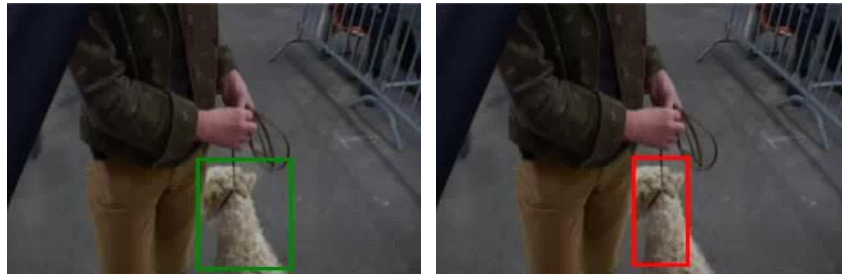
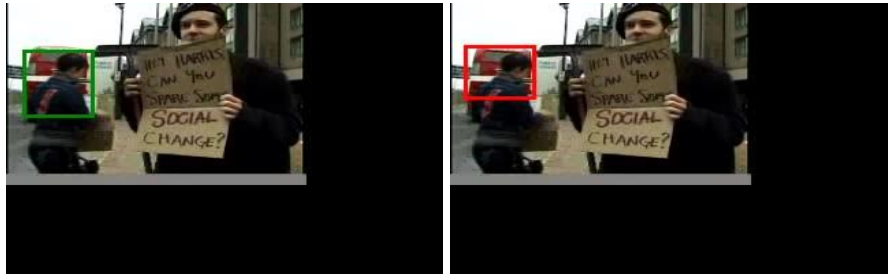
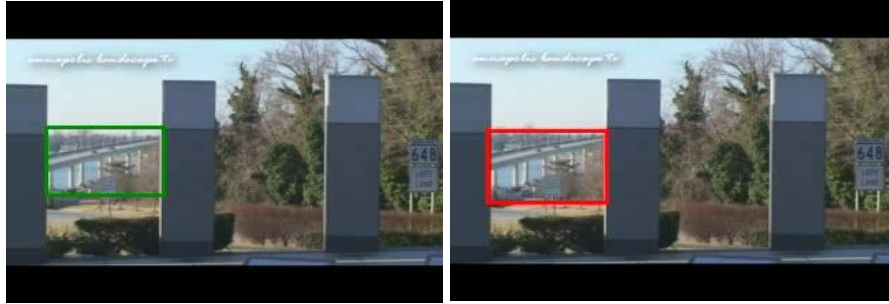


### Spatial localization



□ GT  
□ Sys

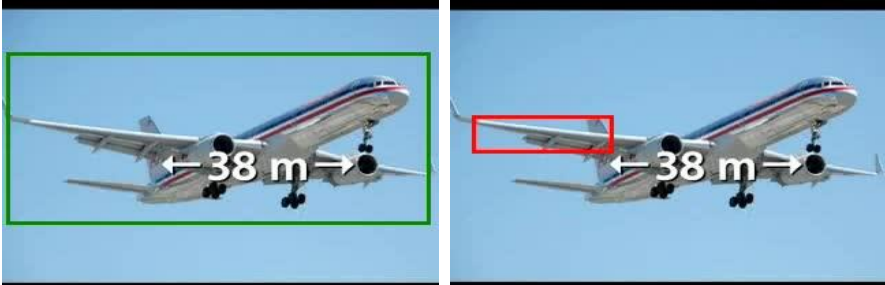
Samples of good localization



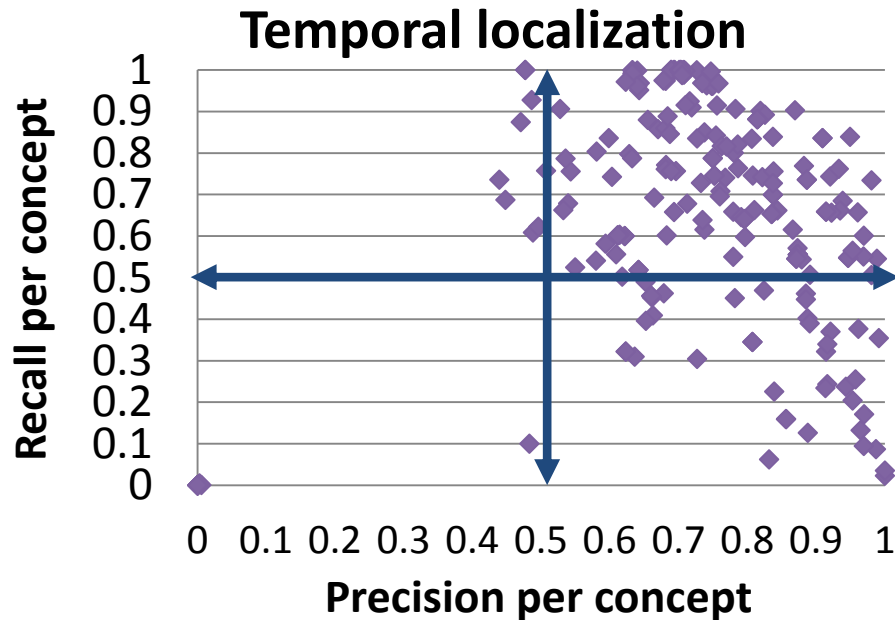


□ GT  
□ Sys

Samples of **less good** localization

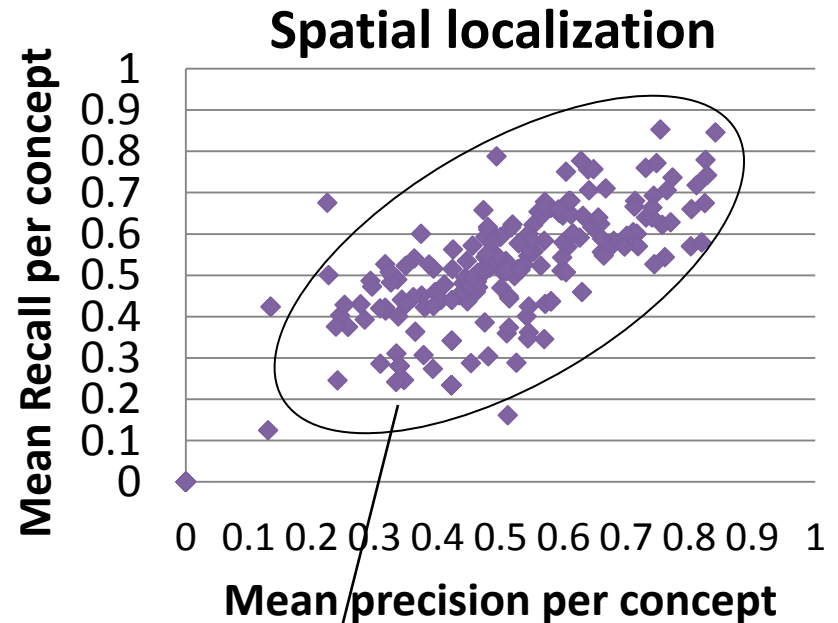


# Results per concept across all runs



Retrieval of target I-frames are much better than 2013 & 2014

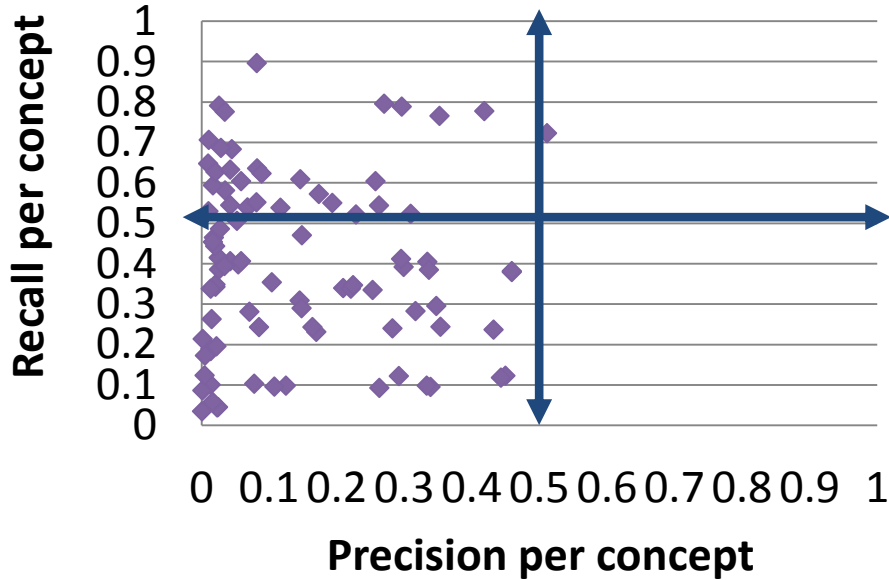
Due to focusing only on TP SIN shots ?



submitted bounding boxes approximate G.T boxes in size with some overlap. Systems are good in finding the real box sizes.

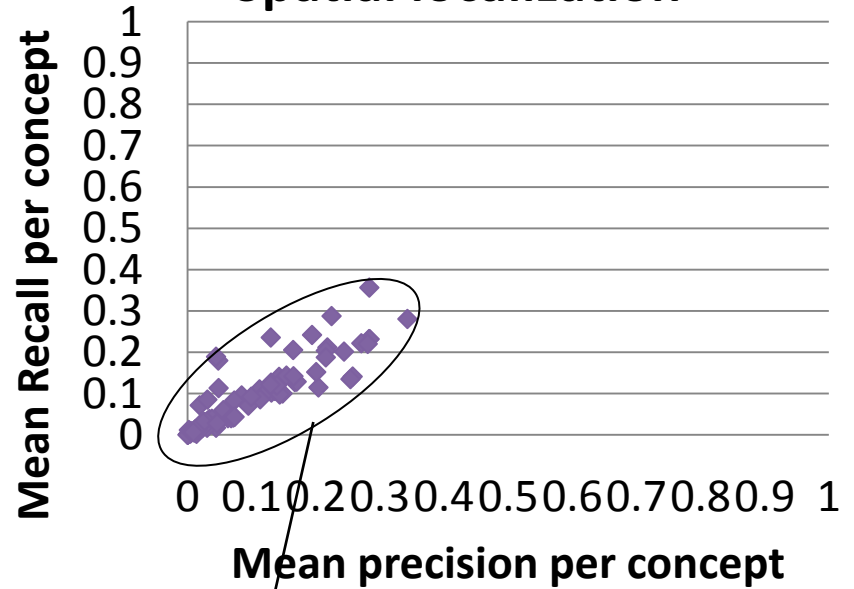
# Results per concept across all teams 2013

## Temporal localization



Majority of systems submitted a lot of non-target I-frames.

## Spatial localization



Most systems submitted bounding boxes  $\approx$  G.T boxes AND overlaps. Systems are good in finding the real box sizes 😊

# Observations/Questions

- Temporal localization is easier than spatial localization.
- The high performance this year was due to:
  - The task was run independently from SIN (systems were given a set of TP shots to localize) made it too easy ?
  - The reuse of the same concepts and availability of previous box annotations ?
  - There is a real new enhancements in the proposed techniques ?
- Should we run the task again in 2016 ?
  - Time for new testing concepts ?
  - IACC.1 and IACC.2 datasets can be reused on new concepts
  - Use testing shots that are combination of TPs + TNs (simulating raw SIN run submission)