VOT2013
IEEE workshop on visual object tracking challenge

The Visual Object Tracking VOT2013 Challenge and Results

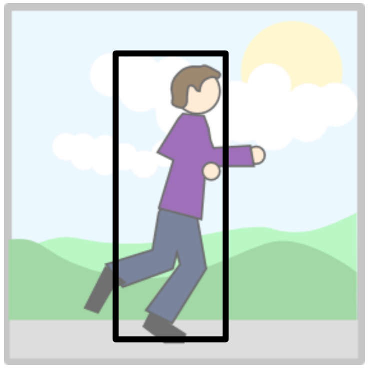
Matej Kristan, Roman Pflugfelder, Aleš Leonardis, Jiri Matas, Fatih Porikli, Luka Čehovin, Georg Nebehay, Gustavo Fernandez, Tomaš Vojir, et al.

AUSTRIAN INSTITUTE OF TECHNOLOGY
University of Ljubljana
Faculty of Computer and Information Science
UNIVERSITY OF BIRMINGHAM
epics
Engineering Proprioception in Computing Systems
NICTA
Australian National University

Outline

1. Scope of the challenge
2. Evaluation system
3. Dataset
4. Performance measures
5. Submitted trackers
6. Experiments and results
7. Summary

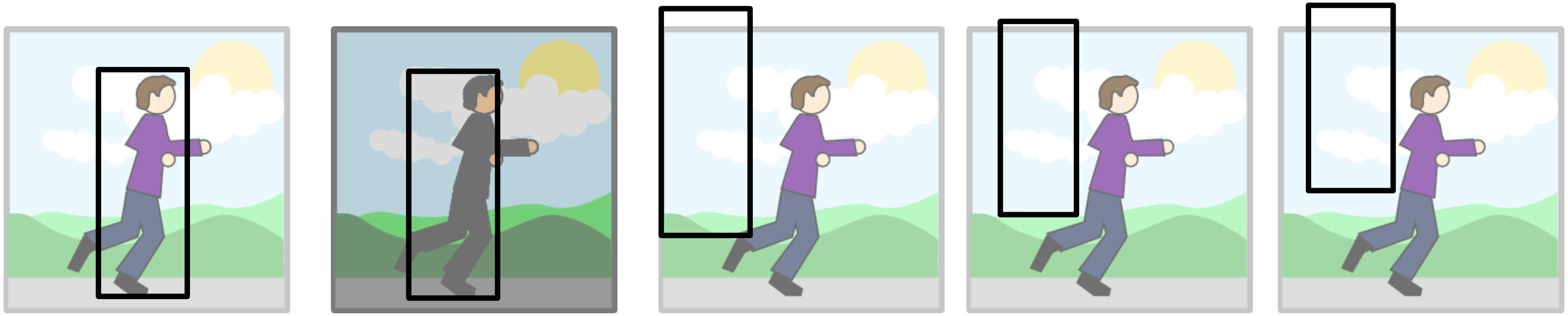
Class of trackers tested



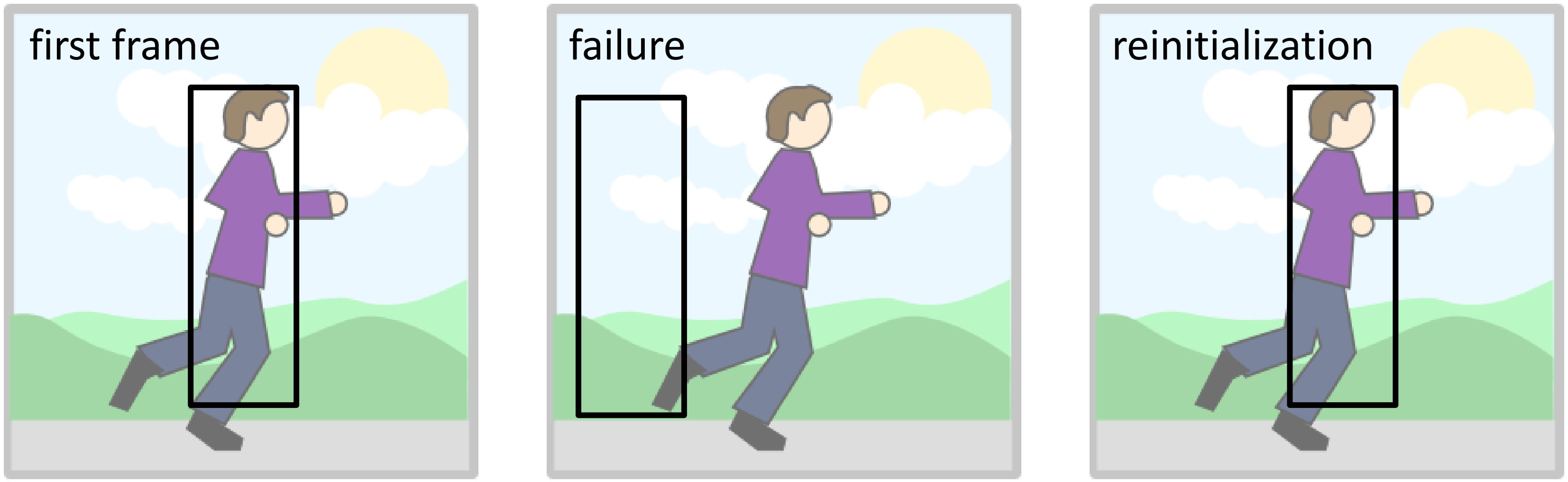
- Single-object, single-camera
- Short-term causal tracking
- Short-term:
  - Trackers performing without re-detection
- Causality:
  - Tracker is not allowed to use any future frames
- No prior knowledge about the target
  - Only a single training example – BBox in the first frame
- Object state encoded by an axis-aligned bounding box

Requirements for tracker implementation

- Would like to use the data fully



- **Renitialize** once the tracker drifts from the object



Requirements for tracker implementation

- Complete reset:
  - Memoryless – reinitalization resets the tracker
  - Tracker is not allowed to use any information obtained before reset, e.g., learnt dynamics, visual model.
- Trackers required to predict a single BB per frame
- Parameters may be set internally, but not by detecting a specific sequence
  - Verified for the top-performing trackers
- A change of parameters was not considered a different tracker

VOT2013 EVALUATION SYSTEM

Evaluation system requirements

- Require an evaluation system that automatically performs a battery of experiments
  - Large number of experiments possible
  - Minimize human error
  - Consistency of the results

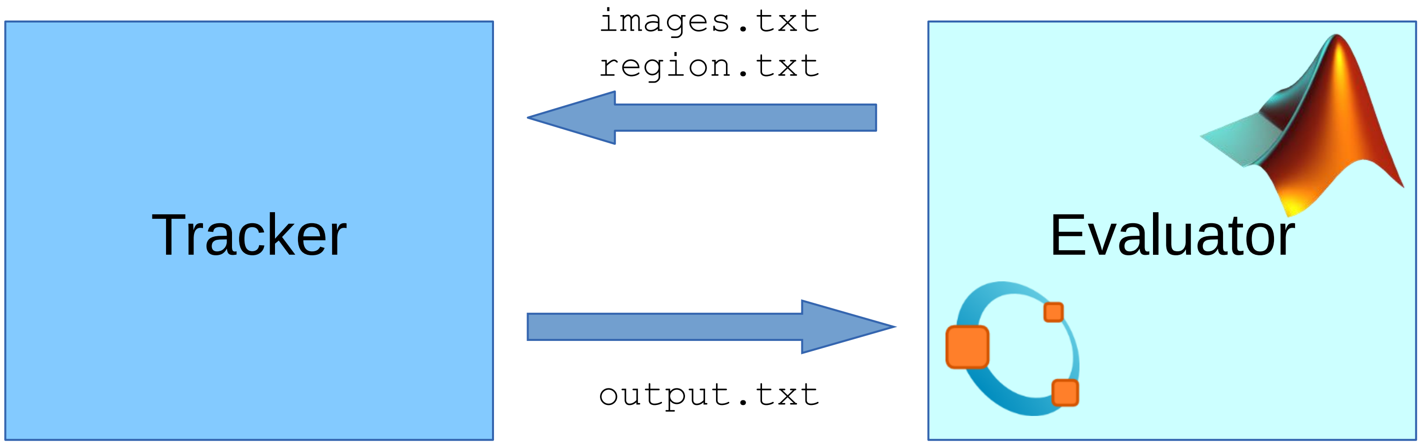
- Requirements
  - Must support multiple platforms
  - Tracker integration not too difficult
  - Must allow reinitialization

Evaluation systems

- ODViS [Jaynes et al., 2002], VIVID [Collins et al., 2005], ViPER [Doermann and Mihalcik 2000]
  - Cannot simply modify for reinitialization
- „Large benchmark experiment“ [Wu et al. CVPR2013]
  - No standardised input-output
  - Integration not straightforward
- Metaanalysis – Evaluation by collecting results from existing publications [Pang et al. ICCV2013]
  - Different approach
  - Not appropriate for recently published trackers

VOT2013 Challenge evaluation kit

- Evaluation kit – download from VOT2013 homepage
- Integration effort minimum

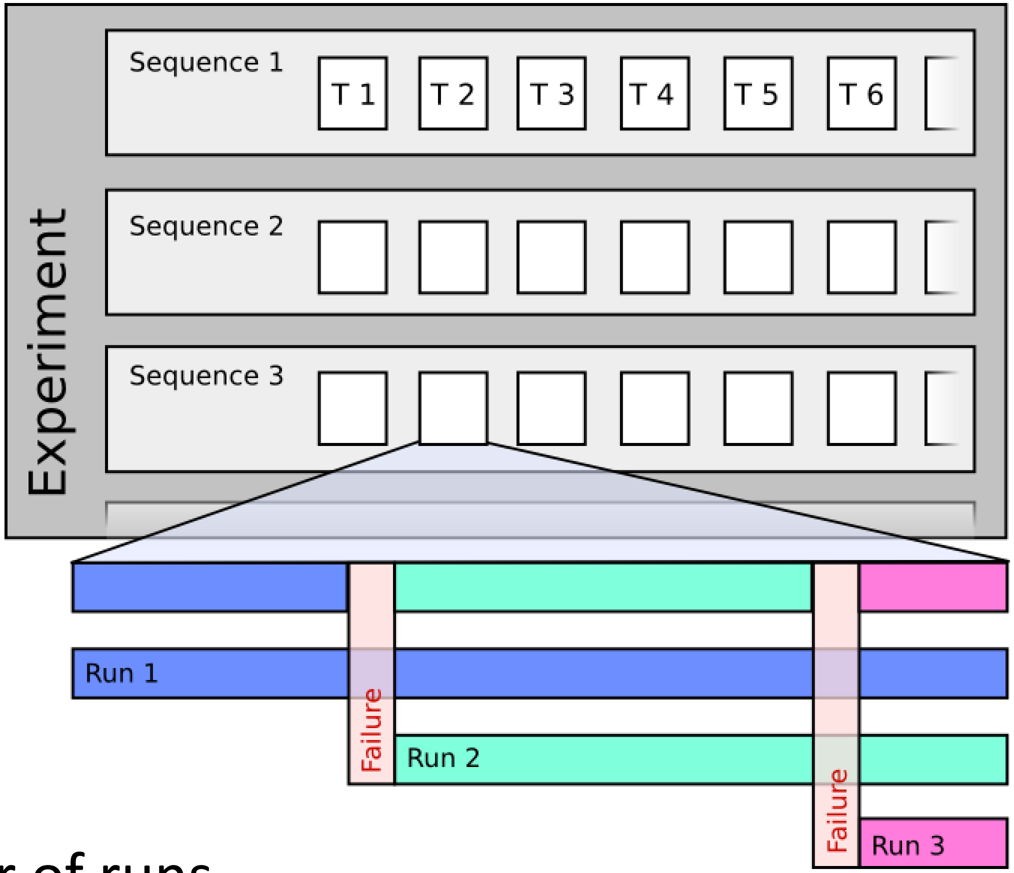


- Runs in Matlab/Octave (multiple platforms)
- Runs the executable (comunication via input parameters)
  - multiple programming languages

https://github.com/vicoslab/vot-toolkit

VOT2013 Challenge evaluation kit

- Pass a sequence + intial BB to tracker (tracks till the end)
- Inspect the output, detect first failure reinitialize from frame $t + \Delta$



*can lead to a large number of runs

VOT2013 DATASET

Dataset: Diverse, not necessarily large

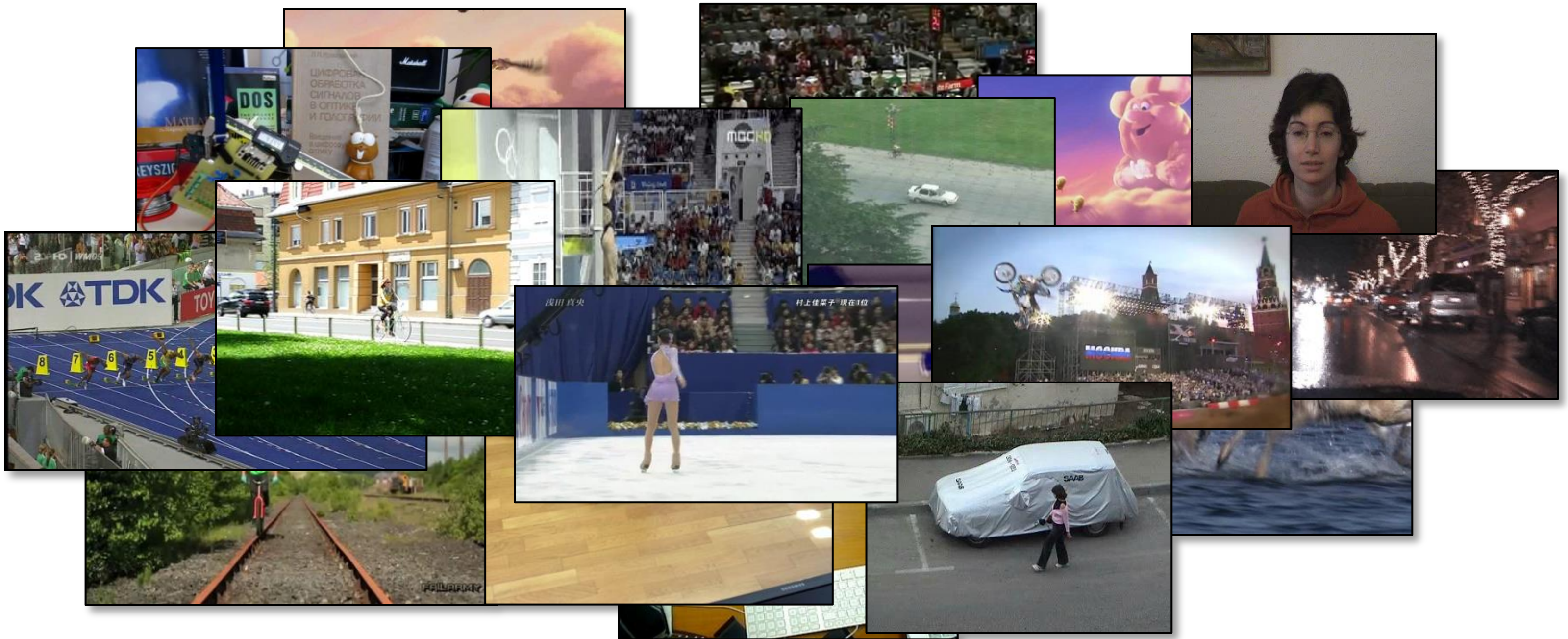
- Lots of datasets: PETS [Young and Ferryman 2005], CAVIAR[1] , i-LIDS[2], ETISEO[3], CVBASE[4], FERET [Phillips et al., 2000], ALOV [Smeulders et al., 2013]

- Diversity in attributes
  - illumination change,
  - dynamic background, object motion, occlusion, etc.
  - camera motion
  - compression artefacts, camera gain, etc.

[1] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1
[2] http://www.homeoffice.gov.uk/science-research/hosdb/i-lids
[3] http://www-sop.inria.fr/orion/ETISEO
[4] http://vision.fe.uni-lj.si/cvbase06/

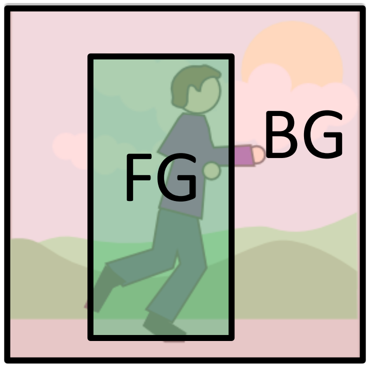
Dataset construction

- Approach:
  - Include various attributes
  - Keep number of sequences low (Time for performing experiments)
- Initially collected a pool of ~60 sequences commonly used in the community



VOT2013 dataset

- Attributes were estimated automatically
  - estimators based on ad hoc heuristics
  - sufficient for sequence selection

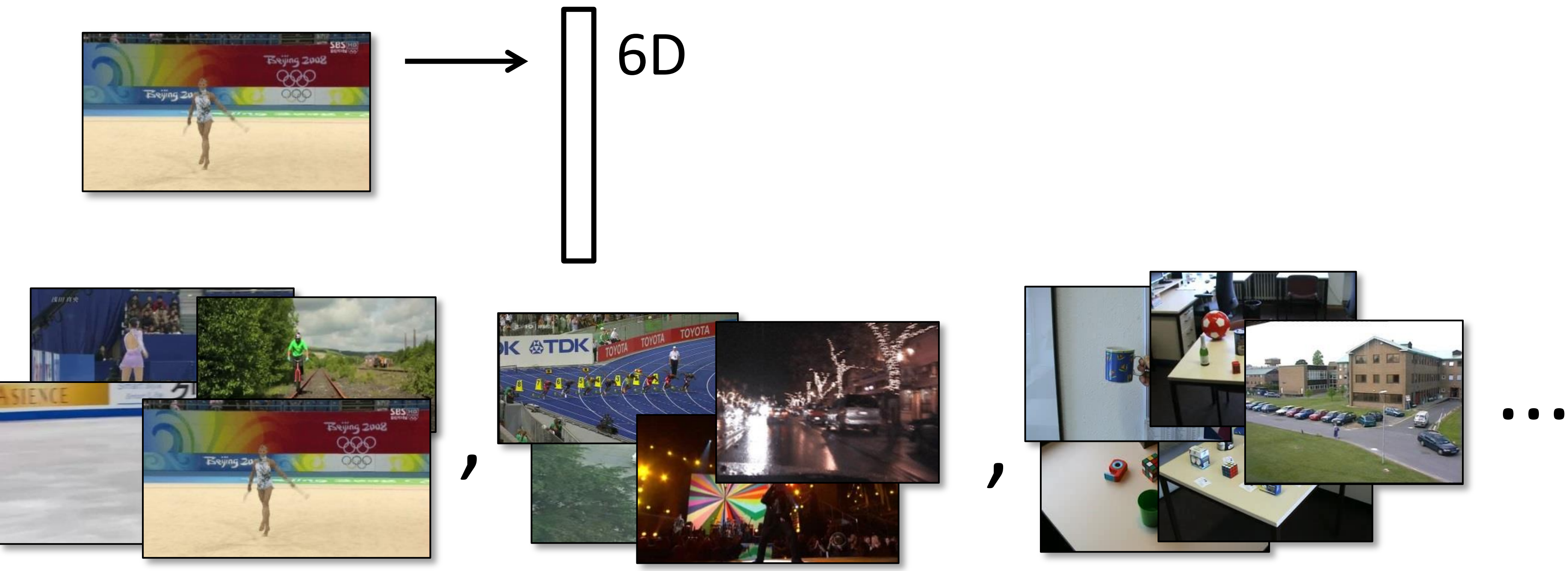


The attributes:

1. Illumination change (difference of min/max FG intensity)
2. Size change (average of sequential BB size difference)
3. Motion (average of sequential BB center difference)
4. Clutter (FG/BG color histogram difference)
5. Camera motion (BG per-pixel differences)
6. Blur (Camera focus measure [Kristan et al., 2006])

VOT2013 dataset

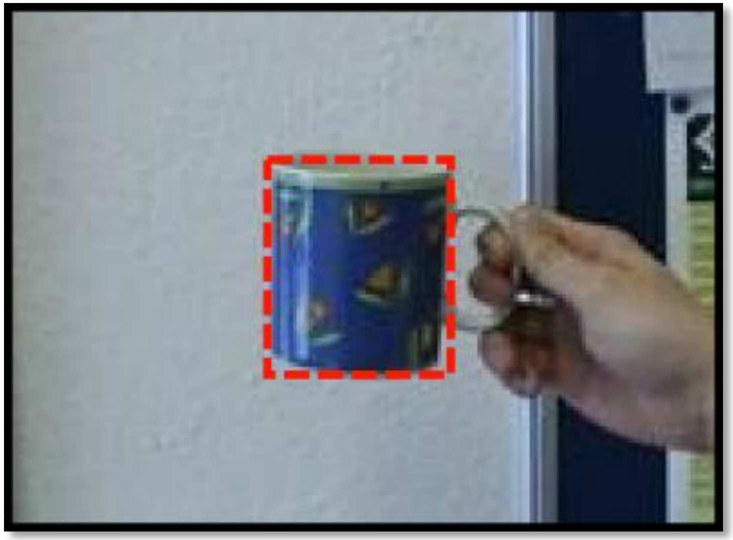
- Sequences clustered into 16 clusters by attributes using Affinity propagation [Frey and Dueck 2007].
- A single video selected from each cluster manually.
  - Make sure that phenomena like occlusion were still well represented.



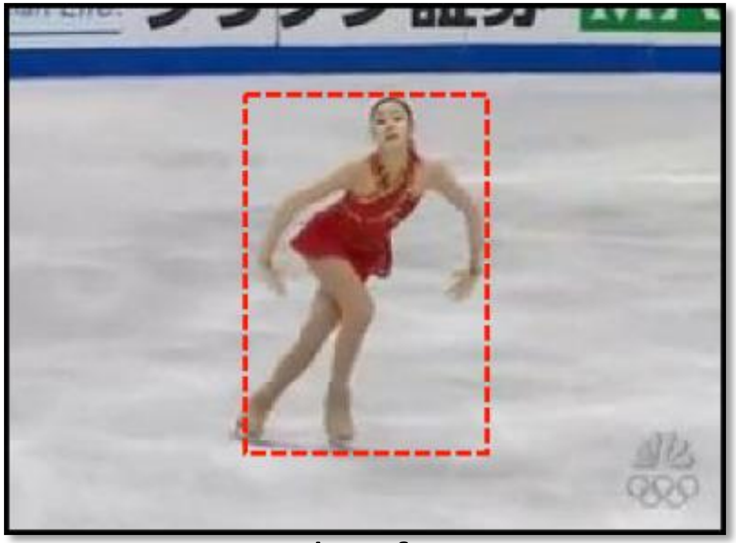
VOT2013 dataset

VOT2013 dataset – object annotation

- Most sequences contained per-frame bounding boxes.
- Annotation by various authors.
- We estimate that >60% of the BB pixels come from the object



example of a BB
for a compact object



example of a BB
for articulated object

Dataset – frame-level annotation

- Common practice: Each sequence annotated by a visual attribute [Dung et al 2010,Wu et al. 2012]
- However, a visual phenomenon does not last over entire sequence

A failure might incorrectly interpreted
as the failure due to occlusion (which happens later on!)

- For a detailed analysis we require per-frame annotations.

VOT2013 dataset – frame annotation

- Manually and semi-manually labeled each frame with visual attributes:
    - i. Occlusion (M)
    - ii. Illumination change (M)
    - iii. Object motion (A)
    - iv. Object size change (A)
    - v. Camera motion (M)
    - vi. Nondegraded (A)

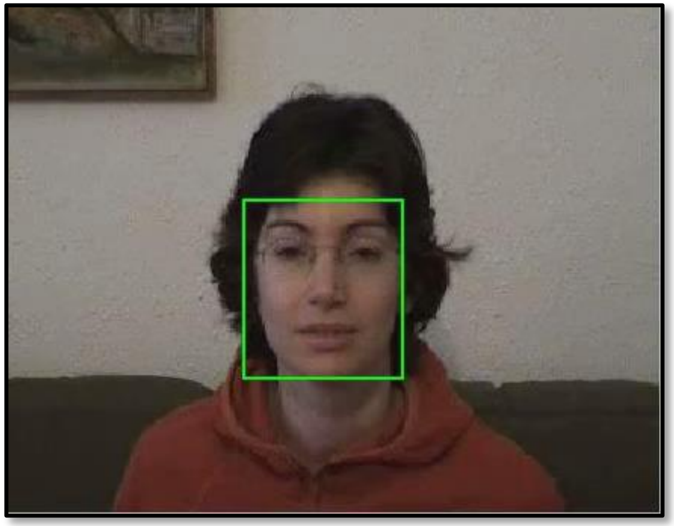
M ... manual annotation, A ... automatic annotation

(i)	**0**	**1**	**1**	**0**
(ii)	**0**	**0**	**0**	**0**
(iii)	**0**	**0**	**0**	**0**
(iv)	**1**	**1**	**1**	**1**
(v)	**0**	**0**	**0**	**0**
(vi)	**0**	**0**	**0**	**0**

VOT2013 dataset – frame annotation

- Example: Occlusion

  All annotations: occlusion



- Example: Illumination change

  All annotations: camera motion, illumination change, motion

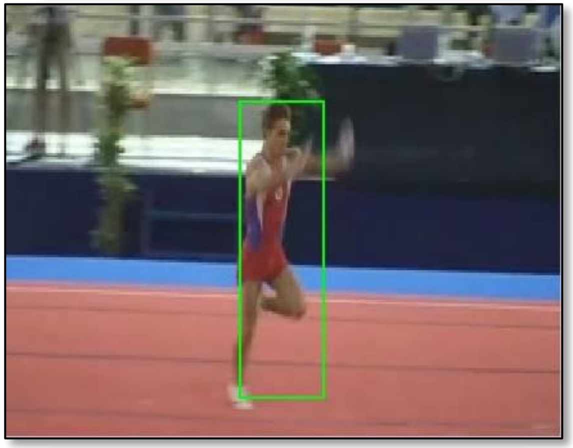
VOT2013 dataset – frame annotation

- Example: Object motion

  All annotations:motion, size

- Example: Object size change

  All annotations: camera motion, motion, size



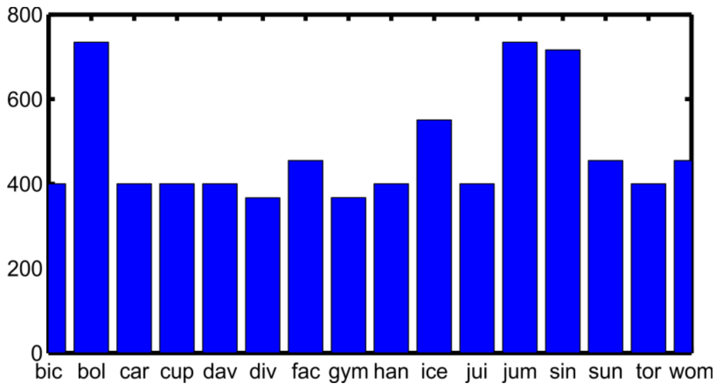
- Example: Camera motion

  All annotations: camera motion, motion

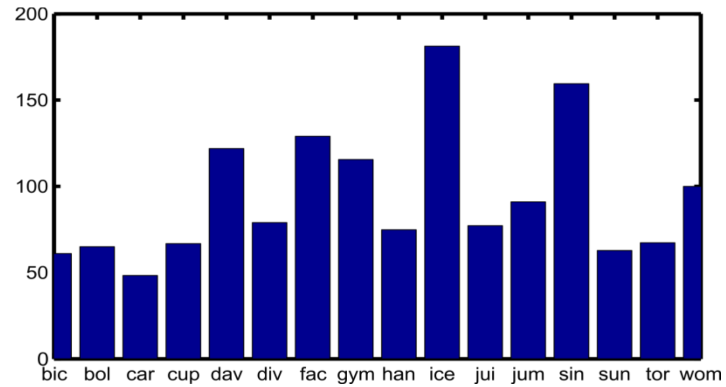
VOT2013 dataset – general stats

- 16 color sequences:

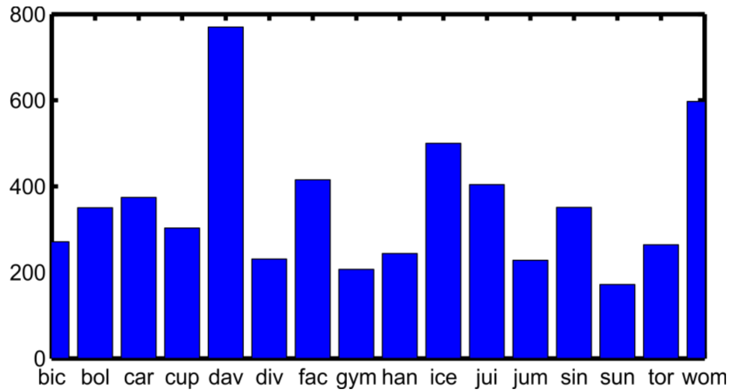
Diagonals of images



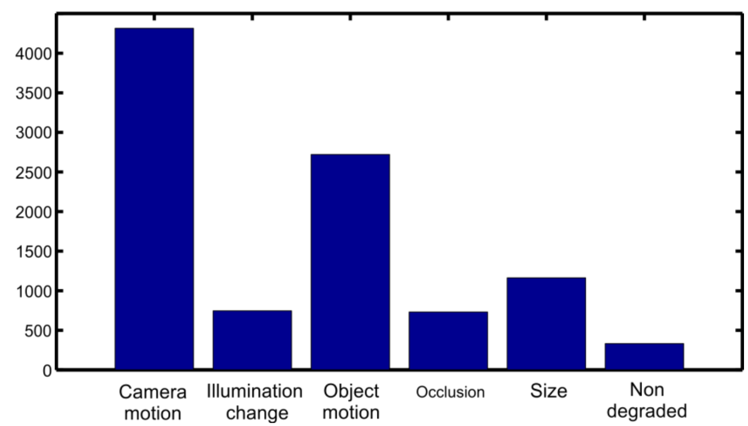
Object bounding box diagonals



sequence length distribution



frames per attribute



EVALUATION METHODOLOGY

Performance measures

- A wealth of performance measures exist
- Basic ones: center distance, region overlap, tracking length, failure rate
- Basic measures offer a straight-forward interpretation
- Combined ones: CoTPS [Nawaz&Cavalaro 2013]
  - Combination of region overlap and tracking length.
- Recent study [Čehovin et al. 2013] has shown that many basic tracking mesures are correlated.
  - Combining correlated measures may introduce bias!

VOT2013 performance measures

- Approach:
  - Interpretability of a measure
  - Select as few as possible to provide clear comparison

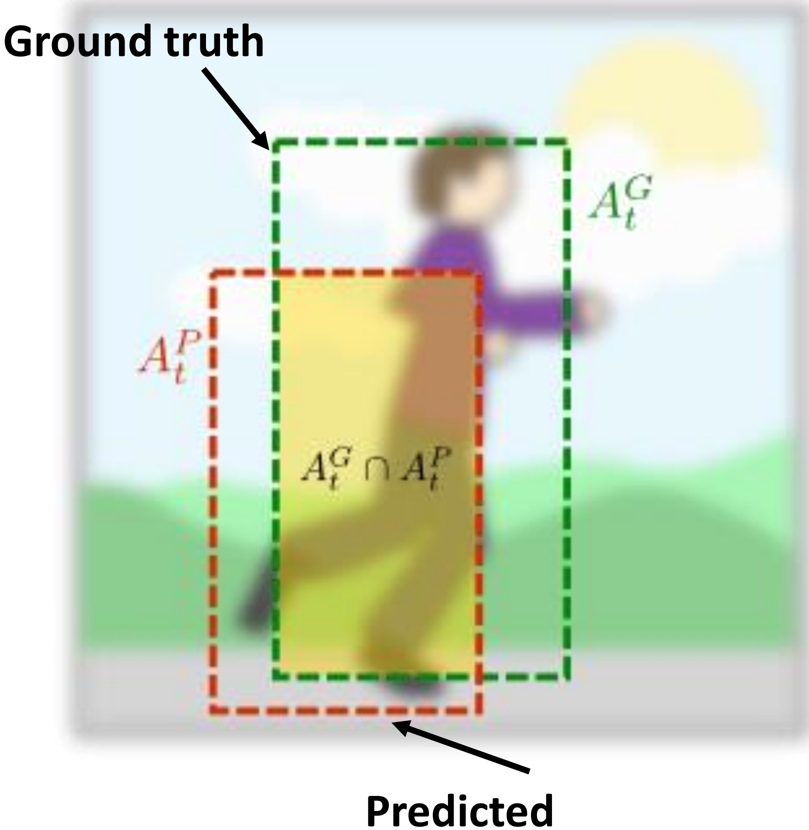
- Based on the recent study[1] we chose two basic weakly-correlated measures:
  - Accuracy
  - Robustness

[1][Čehovin2013] Čehovin, Kristan and Leonardis „Is my tracker new really better than yours?“, Technical Report, ViCoS ,2013 (link)

VOT2013 measures: Accuracy

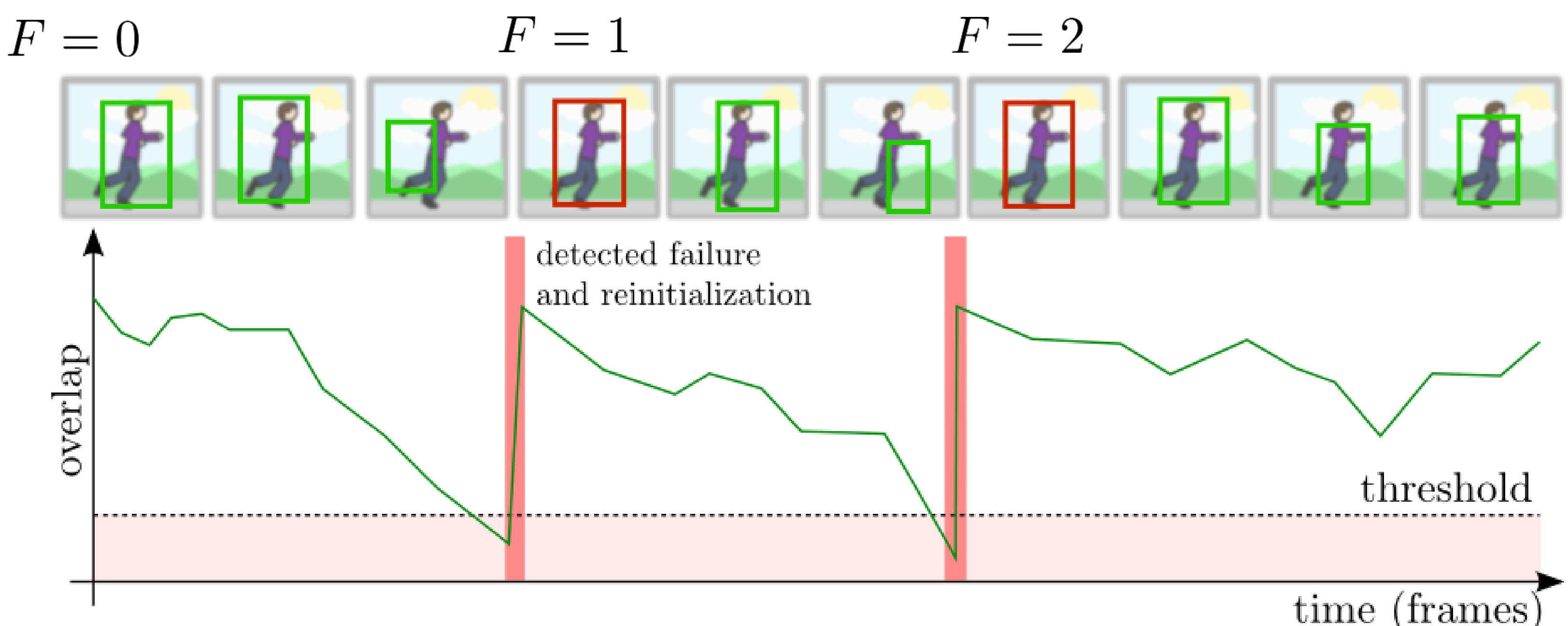
- Overlap between the ground-truth BB and the BB, predicted by a tracker

$$\Phi(\Lambda_G, \Lambda_P) = \left\{ \frac{A_t^G \cap A_t^P}{A_t^G \cup A_t^P} \right\}_{t=1}^{N}$$



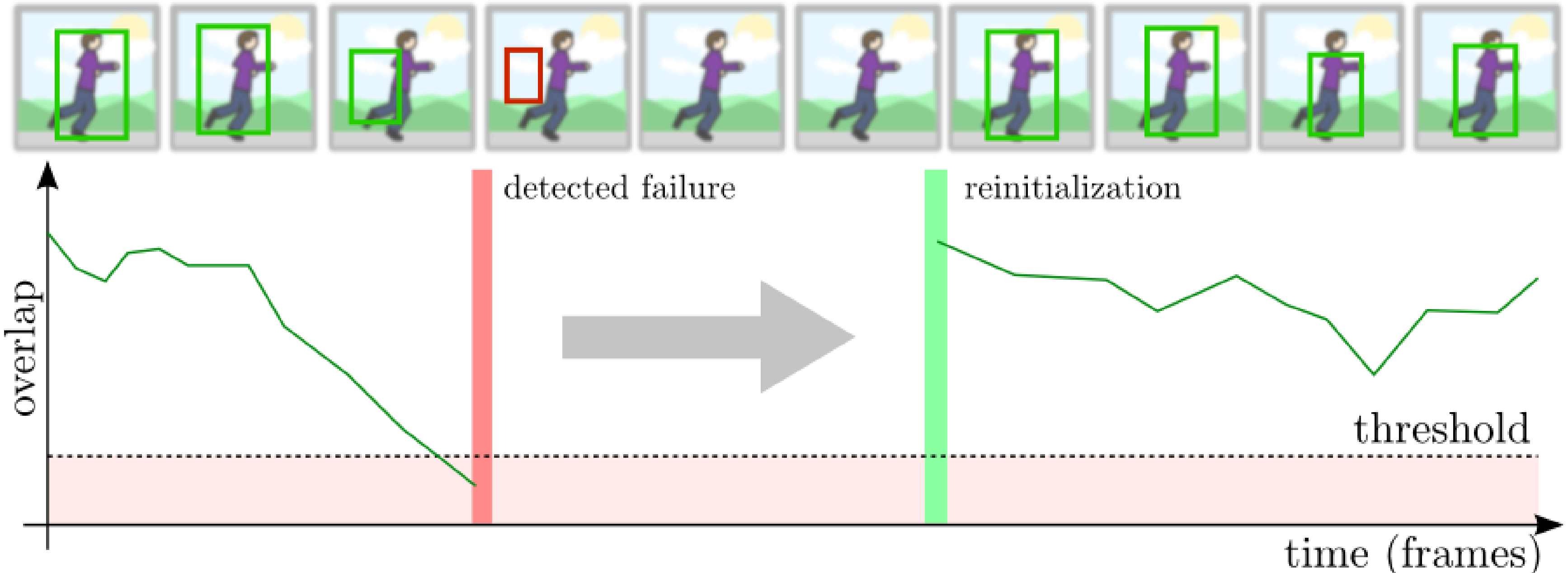
VOT2013 measures: Robustness

- Counts the number of times the tracker failed and had to be reinitialized
- Failure detected when the overlap $\Phi(\Lambda_G, \Lambda_P)$ drops below a threshold



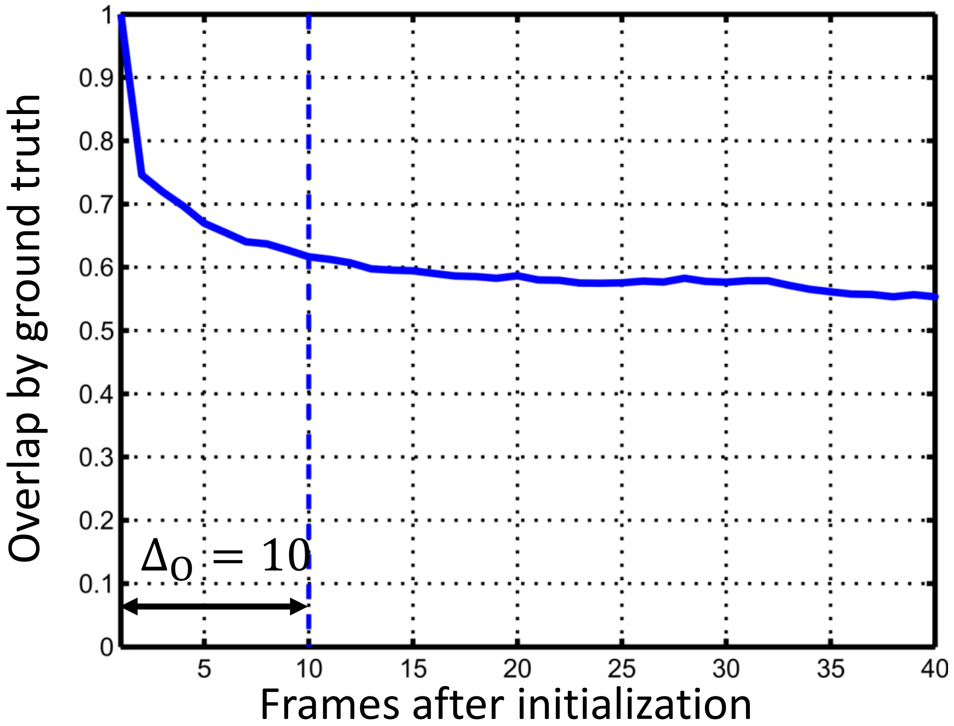
VOT2013 measures: Reinitialization

- If a tracker fails in one frame it will likely fail again if reinitialized in the next frame.
- To avoid this correlation we reinitialize the tracker $\Delta_{\mathrm{F}} = 5$ frames after the failure.
- Δ_{F} determined experimentally on a separate dataset



VOT2013 measures: Reinitialization

- Overlaps immediately after reinitialization biased toward higher values.
- Burn-in period required to reduce initialization bias
- The curve flattens at $\Delta_O = 10$ frames



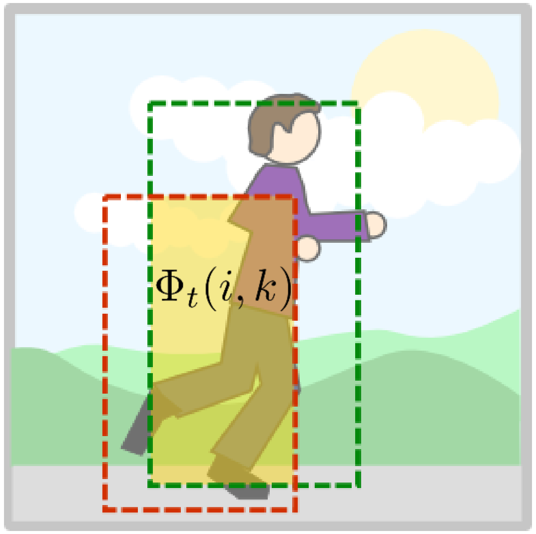
Preliminary test:

- Initialize many trackers
- Record overap
- Average at each frame

VOT2013 measures: Multiple runs

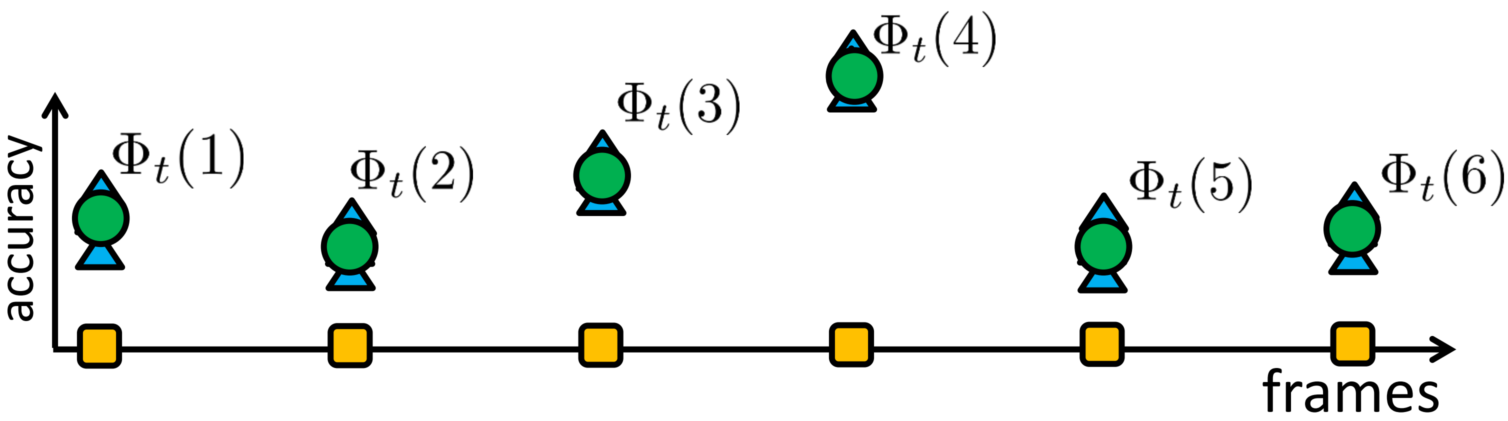
- Measures averaged over multiple runs

  $\Phi_t(i,k)$ ... accuracy of *i*-th tracker at frame *t* at repetition *k*.



- Per-frame averaged accuracy

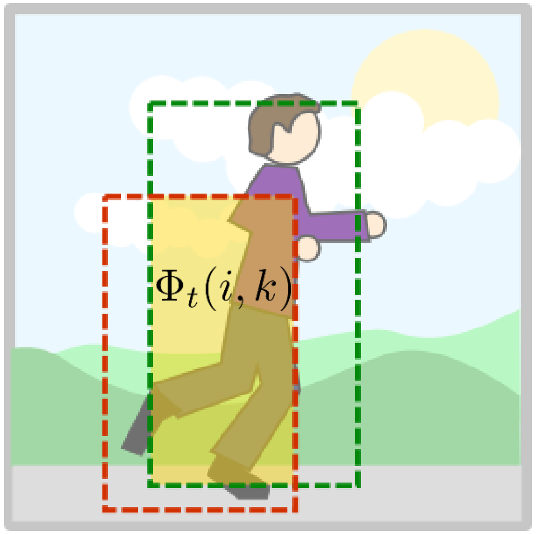
$$\Phi_t(i) = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \Phi_t(i,k)$$



VOT2013 measures: Multiple runs

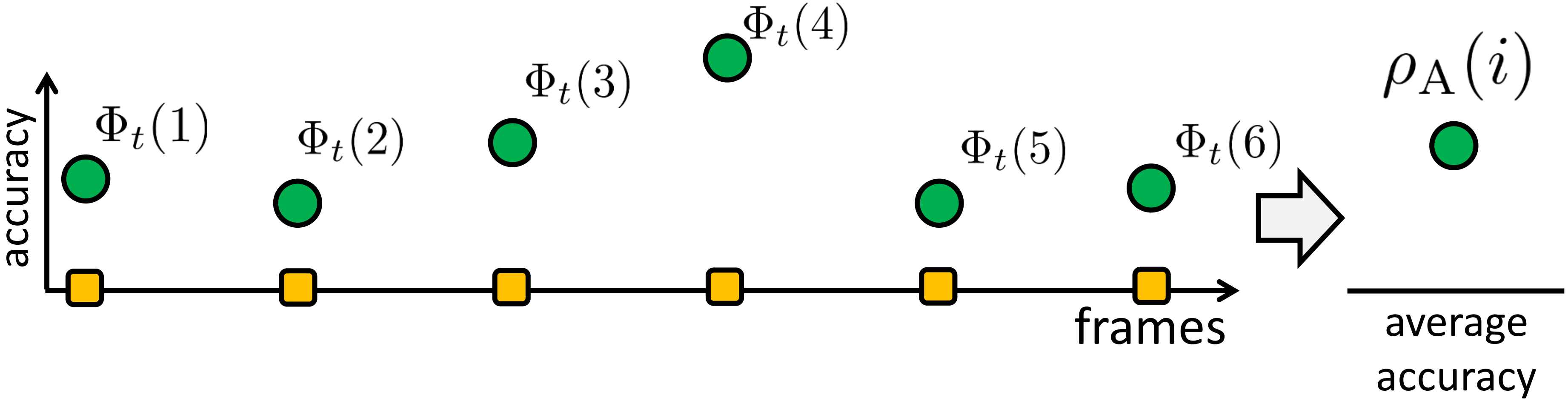
- Average accuracy at frame *t*

$$\Phi_t(i) = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \Phi_t(i,k)$$



- Average accuracy over sequence

$$\rho_{\mathrm{A}}(i) = \frac{1}{N_{\mathrm{valid}}} \sum_{j=1}^{N_{\mathrm{valid}}} \Phi_j(i)$$



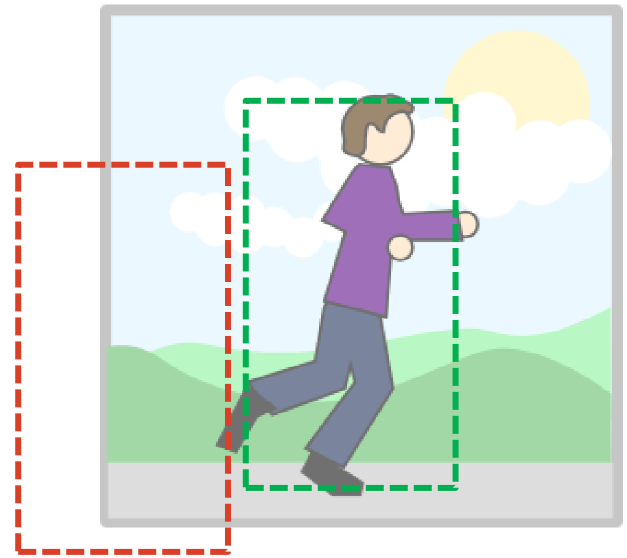
VOT2013 measures: Multiple runs

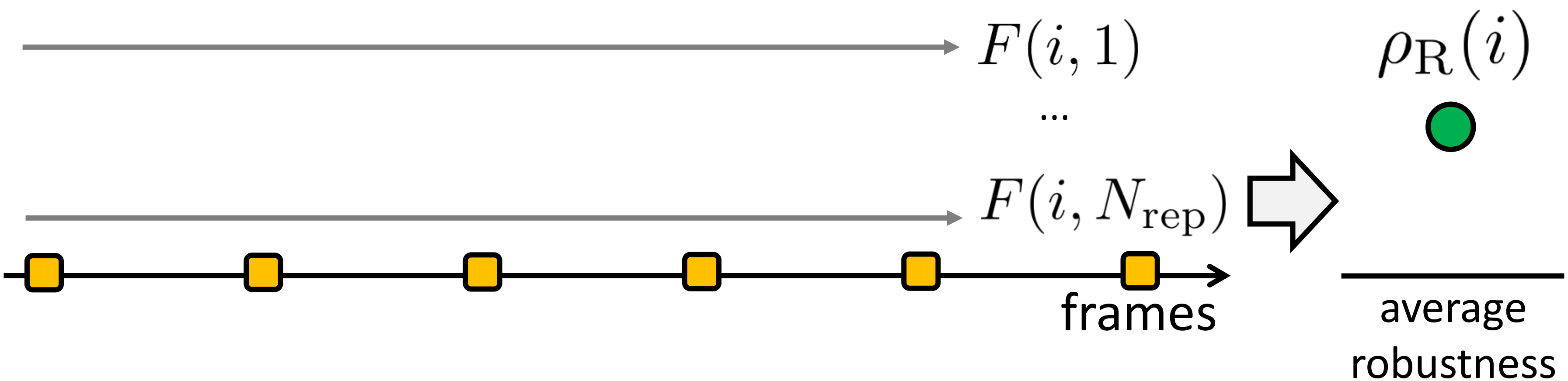
- Multiple measurements of robustness (#failures)

$F(i,k)$ ... number of failures of *i*-th tracker at repetition *k*.

- Average robustness per sequence

$$\rho_R(i) = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} F(i,k)$$

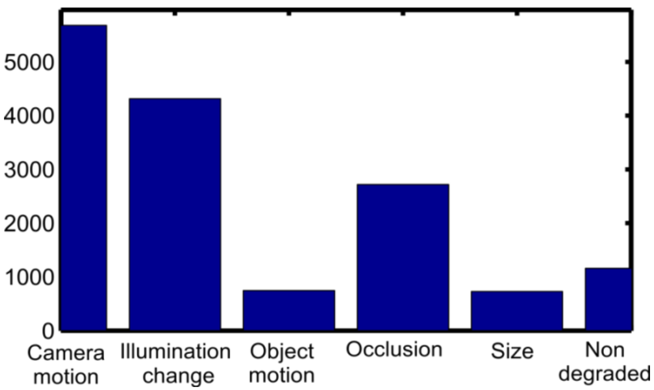




Measures: Attribute weighting

- Attribute subset: In all sequences consider only frames that correspond to a particular attribute.

- Compute the average performance measures ρ_A, ρ_R for each attribute subset.



attribute a1 seq. $\Rightarrow [\rho_A(i, a_1), \rho_R(i, a_1)]$

attribute a2 seq. $\Rightarrow [\rho_A(i, a_2), \rho_R(i, a_2)]$

...

attribute a6 seq. $\Rightarrow [\rho_A(i, a_6), \rho_R(i, a_6)]$

Primary performance measure: overall rank r(.)

1. Rank trackers for each performance measure separately on each attribute subset.

   $r(i, a, m)$ ... rank of a tracker *i* on attribute subset *a*, evaluated for perfomance measure *m*.

2. Average ranking over the attributes

   $$r(i, m) = \frac{1}{N_{\mathrm{att}}} \sum_{a=1}^{N_{\mathrm{att}}} r(i, a, m)$$

3. Giving equal weight to each performace measure we average the two corresponding rankings

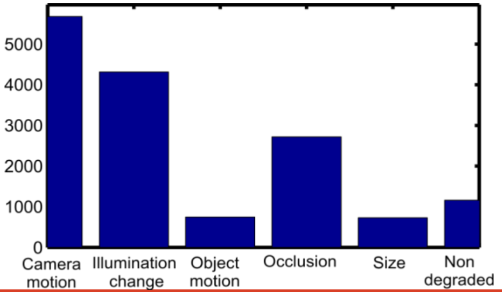
   $$r(i) = \frac{1}{2} \sum_{m \in \{A, R\}} r(i, m)$$

Notes on overall rank

Performance on attribute a_1 subset :

Tracker *i*	T_1	T_2	T_3	T_4
$\rho_A(i, a_1)$	1	0.1	0.5	0.7
$\rho_R(i, a_1)$	0.1	7	10	5
$r(i, a_1, A)$	1	4	3	2
$r(i, a_1, R)$	1	3	4	2

- Ranking-based methodology akin to [Goyette et al. 2012]
- Different frames effectively have a different weight – eg., may have multiple attributes.
- Frequency of attributes is uneven
- Each attribute equally important



Tracker rank equality

- Several trackers may perform equally well and should be assigned an equal rank

do not perform equally well

Tracker *i*	T_1	T_2	T_3	T_4
$\tilde{r}(i, a_1, A)$	1	2	3	4

perform equally well

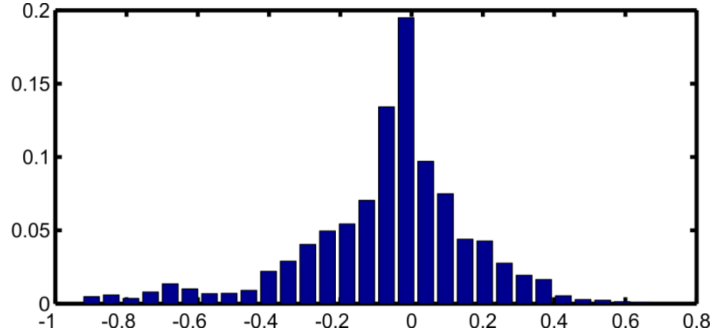
perform equally well

- „Statistical“ equality as defined here is not transitive!
- Modify the ranks by averaging ranks of equivalent trackers

Tracker *i*	T_1	T_2	T_3	T_4
$r(i, a_1, A)$	1.5	2	2.5	4

Statistical equivalence in accuracy

- Per-frame measure available for each tracker.
- Apply a paired test to determine the statistical significance of the differences in accuracy.
- Typically T-test is applied, but assumes a Normal pdf.



- Gaussian assumption might be violated (Anderson-Darling test)
- A nonparametric test for Accuracy:
  - Wilcoxon signed-rank test as in [Demšar IJMLR2006]
  - Tests H_0 that the differences come from a pdf with a zero median

Statistical equivalence in robustness

- Multiple per-sequence measures

attribute a1 seq. $\rightarrow$ For each tracker:
A single robustness measurement per experiment repetition.

- These cannot be paired
- Apply the Wilcoxon Rank-Sum (Mann-Whitney U-test) instead.
  - Two-sided rank sum test of the H_0 that robustness values of T_1 and T_2 are independent samples from pdf with equal medians.

CHALLENGE PARTICIPATION AND SUBMITTED TRACKERS

VOT2013 Challenge: participation

- Authors downloaded
  - The evaluation kit
  - Dataset
- Integrated their tracker into the evaluation kit
- Predefined set of experiments automatically performed
- Participated by submitting the results outputted by the evaluation kit to the VOT2013 challenge.
  - Note: Self-evaluation (experiments run by the authors!)
- Participants were also offered to submit the binaries and/or source code for VOT2013 committee verification of the results

Submitted trackers: 27

19 entries from various authors + 8 baselines contributed by the VOT2013 committee = 27 trackers.

AIF	Chen et al.	VOT, 2013
ASAM	Bozorgtabar and Goecke	?
CACTuS-FL	Wong et al.	IVCNZ, 2010
CCMS	Vojir and Matas	/
CT	Zhang et. al.	ECCV, 2012
DFT	Sevilla-Lara and Learned-Miller	CVPR, 2012
EDFT	Felsberg	VOT, 2013
FoT	Vojir and Matas	CVWW, 2011
HT	Godec et. al.	CVIU, 2013
IVT	Ross et. al.	IJCV, 2008
LGT++	Xiao et. al.	VOT, 2013
LGT	Cehovin et. al.	TPAMI, 2013
LT-FLO	Lebeda et. al.	VOT, 2013
GSDT	Gao et. al.	VOT, 2013

Matrioska	Maresca and Petrosino	ICIAP, 2013
Meanshift	Comaniciu et. al.	TPAMI, 2003
MIL	Babenko et. al.	TPAMI, 2011
MORP	Kraimer	/
ORIA	Wu et. al.	CVPR, 2012
PJS-S	Zarezade et. al.	ArXiv, 2013
PLT	Heng et. al.	/
RDET	Salaheledin et. al.	VOT, 2013
SCTT	Li and Zhu	/
STMT	Poullot and Satoh	/
Struck	Hare et. al.	ICCV, 2011
SwATrack	Lim et. al.	IAPR MVA, 2013
TLD	Kalal et. al.	TPAMI, 2012

Submitted trackers rough categorization

Very diverse set of entries:

- Background-subtraction-based
  (MORP, STMT)
- Optical-flow/motion -based
  (FoT, TLD, SwATrack)
- Key-point-based
  (SCTT, Matrioska)
- Complex appearence-model-based
  (IVT, MS, CCMS, DFT, EDFT, AIF,CactusFl, PJS-S, SwATrack)
- Discriminative models – single part
  (MIL, STRUCK, PLT, CT, RDET, ORIA, ASAM, GSDT)
- Part-based models
  (HT, LGT, LGT++, LT-FLO, TLD)

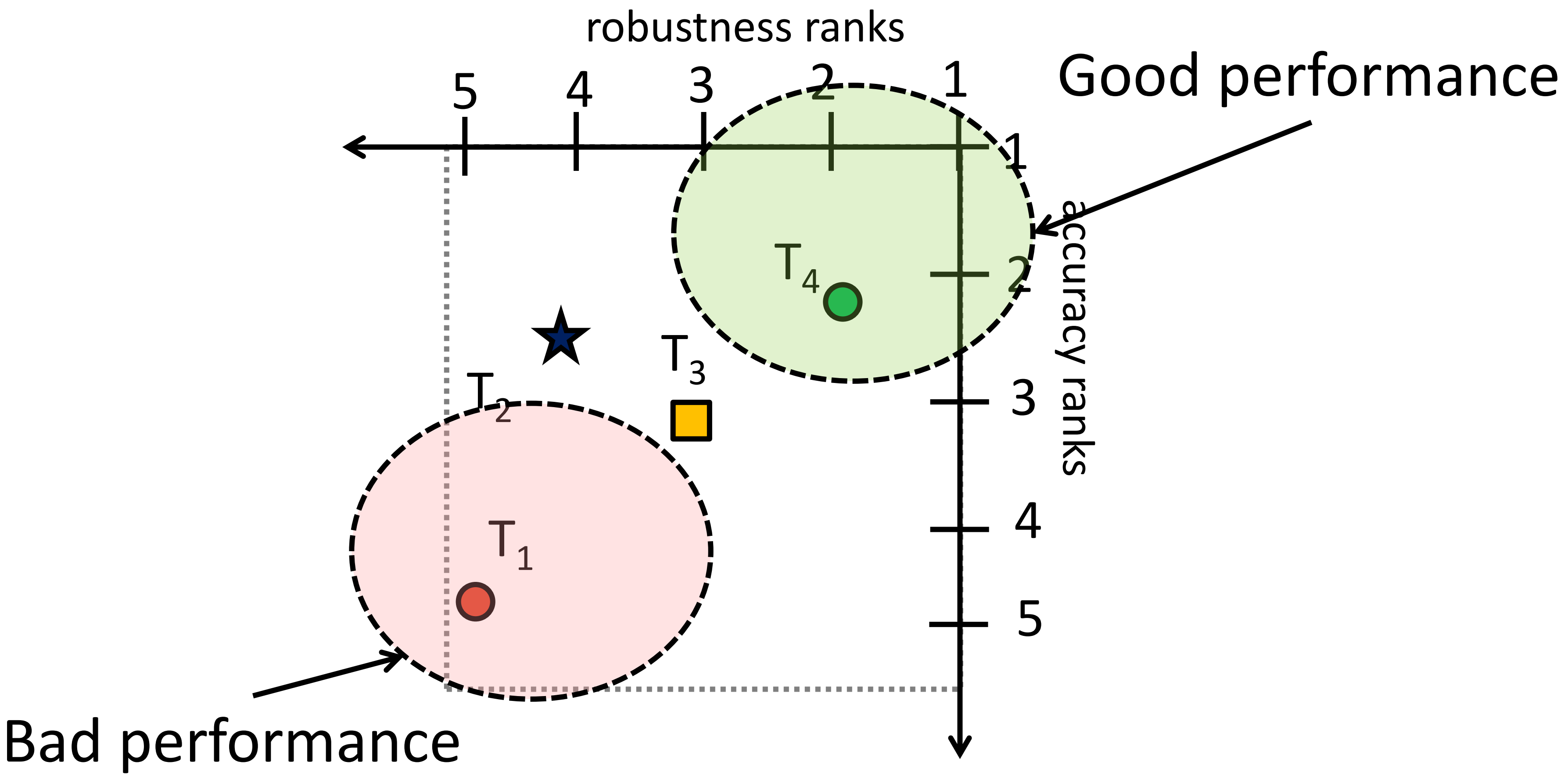
EXPERIMENTS AND RESULTS

VOT2013 Experiments

- Experiment 1– Baseline:
  - All sequences, initialization on ground truth BBs
- Experiment 2 – Noise:
  - Experiment 1 with noisy initialization
  - Perturbations in position and size by drawing uniformly from 10% of the bounding box size.
- Experiment 3 – Grayscale:
  - Experiment 1 with sequences changed to grayscale
- Each tracker run 15 times on each sequence to obtain a better statistic on its performance.
- Reinitialization threshold was 0.

Visualizing the results

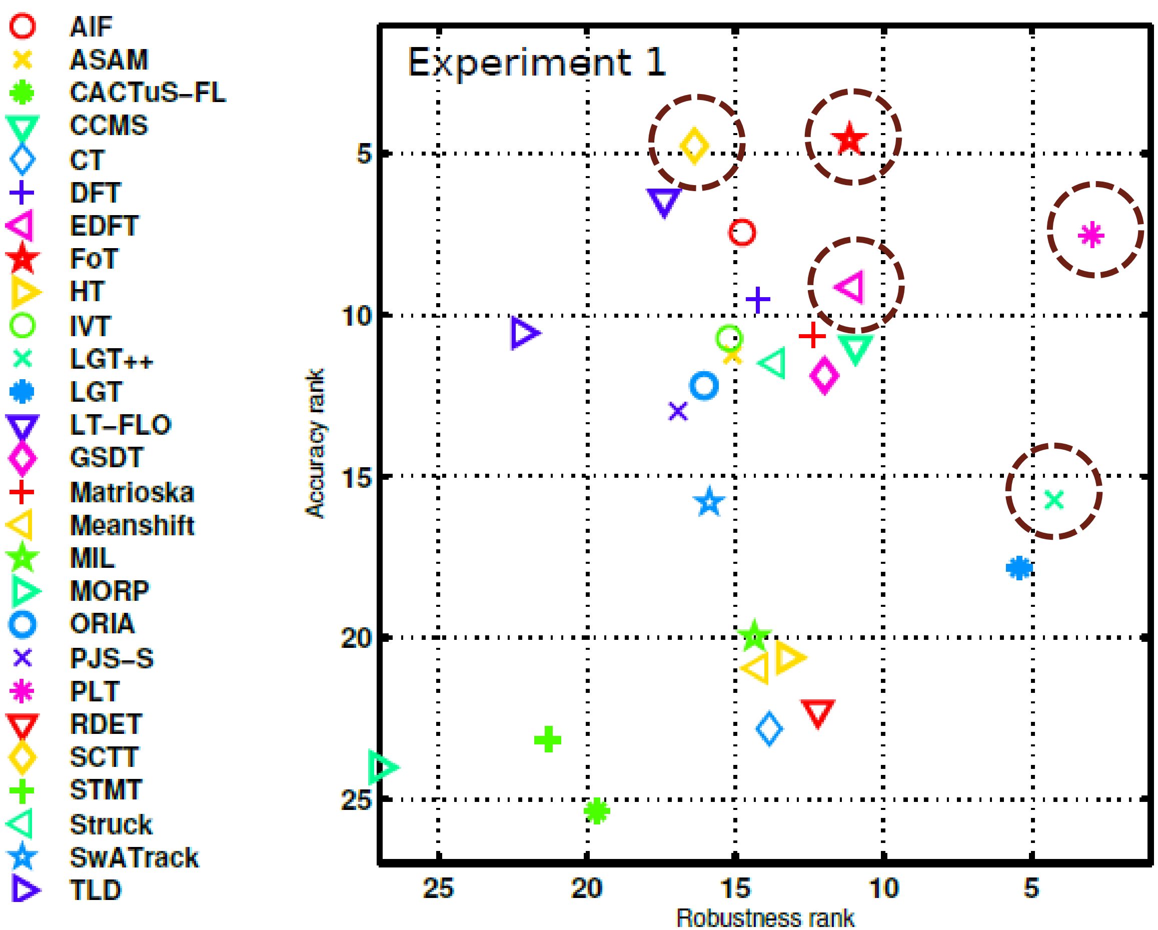
- A-R rank plots inspired by [Čehovin et al. 2013]
  - Each tracker is a single point in the rank space



Results: Experiment 1 (Baseline)

Top performing trackers:

- PLT, FoT, LGT++, EDFT, SCTT



	Experiment 1		
	R_A	R_R	R
PLT*	7.51	3.00	5.26
FoT*	4.56	11.15	7.85
EDFT*	9.14	11.04	10.09
LGT++*	15.73	4.25	9.99
LT-FLO	6.40	17.40	11.90
GSDT	11.87	11.99	11.93
SCTT	4.75	16.38	10.56
CCMS*	10.97	10.95	10.96
LGT*	17.83	5.42	11.62
Matrioska	10.62	12.40	11.51
AIF	7.44	14.77	11.11
Struck*	11.49	13.66	12.58
DFT	9.53	14.24	11.89
IVT*	10.72	15.20	12.96
ORIA*	12.19	16.05	14.12
PJS-S	12.98	16.93	14.96
TLD*	10.55	22.21	16.38
MIL*	19.97	14.35	17.16
RDET	22.25	12.22	17.23
HT*	20.62	13.27	16.95
CT*	22.83	13.86	18.35
Meanshift*	20.95	14.23	17.59
SwATrack	15.81	15.88	15.84
STMT	23.17	21.31	22.24
CACTuS-FL	25.39	19.67	22.53
ASAM	11.23	15.09	13.16
MORP	24.03	27.00	25.51

Results: Experiment 1 (Baseline)

- PLT: single-scale, detection-based tracker that applies online structural SVM on color, grayscale and grayscale derivatives.
- Presentation at: 10:55

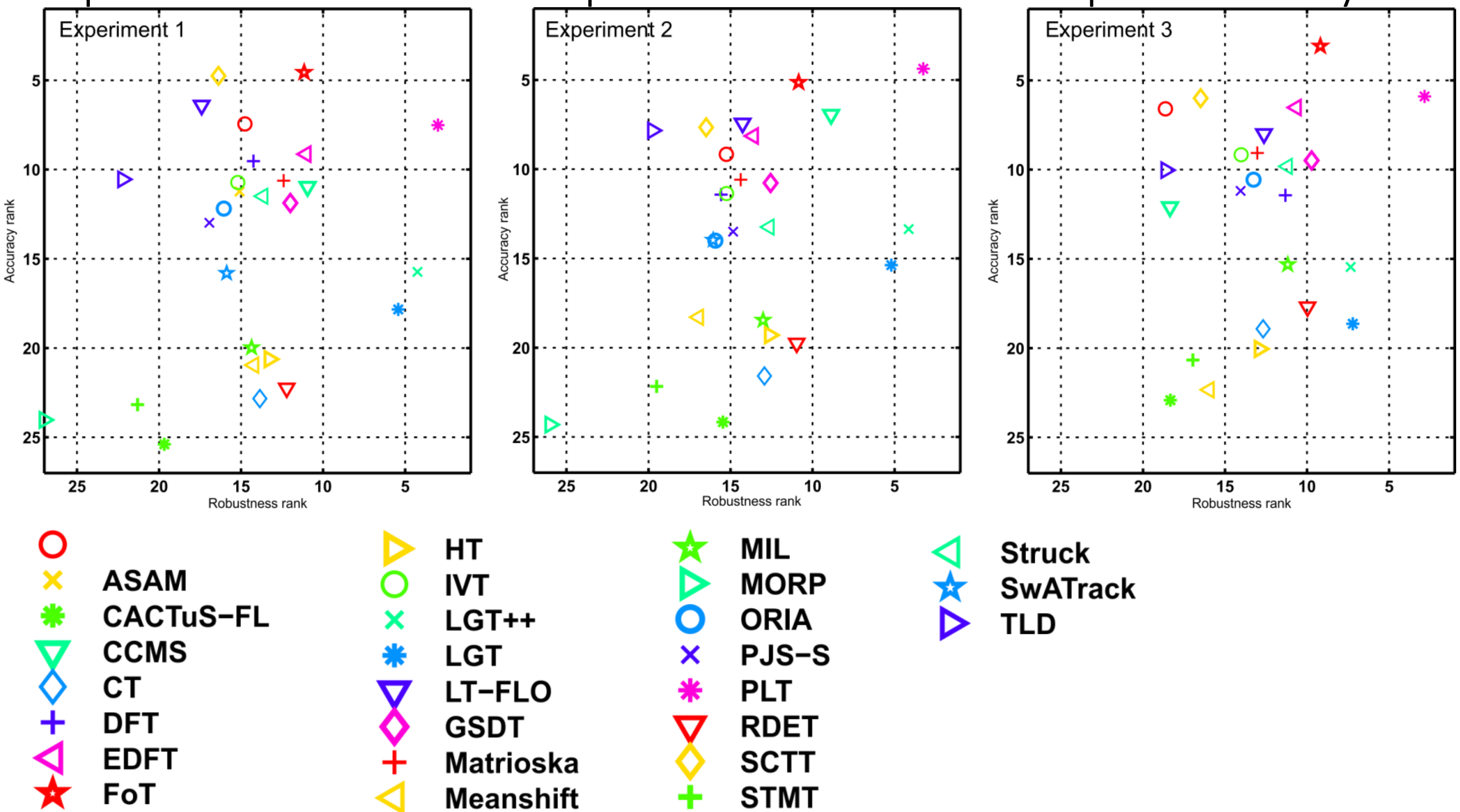
Tracker	Scale adapt.	Dynamic model	Global vis. mod.	Localization
PLT	no	no	no	determinist.
FoT	yes	no	no	determinist.
LGT++	yes	yes	no	stochastic
EDFT	no	yes	yes	determinist.
SCTT	yes	no	no	stochastic

	Experiment 1		
	R_A	R_R	R
PLT*	7.51	3.00	5.26
FoT*	4.56	11.15	7.85
EDFT*	9.14	11.04	10.09
LGT++*	15.73	4.25	9.99
LT-FLO	6.40	17.40	11.90
GSDT	11.87	11.99	11.93
SCTT	4.75	16.38	10.56
CCMS*	10.97	10.95	10.96
LGT*	17.83	5.42	11.62
Matrioska	10.62	12.40	11.51
AIF	7.44	14.77	11.11
Struck*	11.49	13.66	12.58
DFT	9.53	14.24	11.89
IVT*	10.72	15.20	12.96
ORIA*	12.19	16.05	14.12
PJS-S	12.98	16.93	14.96
TLD*	10.55	22.21	16.38
MIL*	19.97	14.35	17.16
RDET	22.25	12.22	17.23
HT*	20.62	13.27	16.95
CT*	22.83	13.86	18.35
Meanshift*	20.95	14.23	17.59
SwATrack	15.81	15.88	15.84
STMT	23.17	21.31	22.24
CACTuS-FL	25.39	19.67	22.53
ASAM	11.23	15.09	13.16
MORP	24.03	27.00	25.51

Results: Experiments 1,2,3

- Considering all 3 experiments:
  PLT, FoT, EDFT, LGT++, LT-FLO

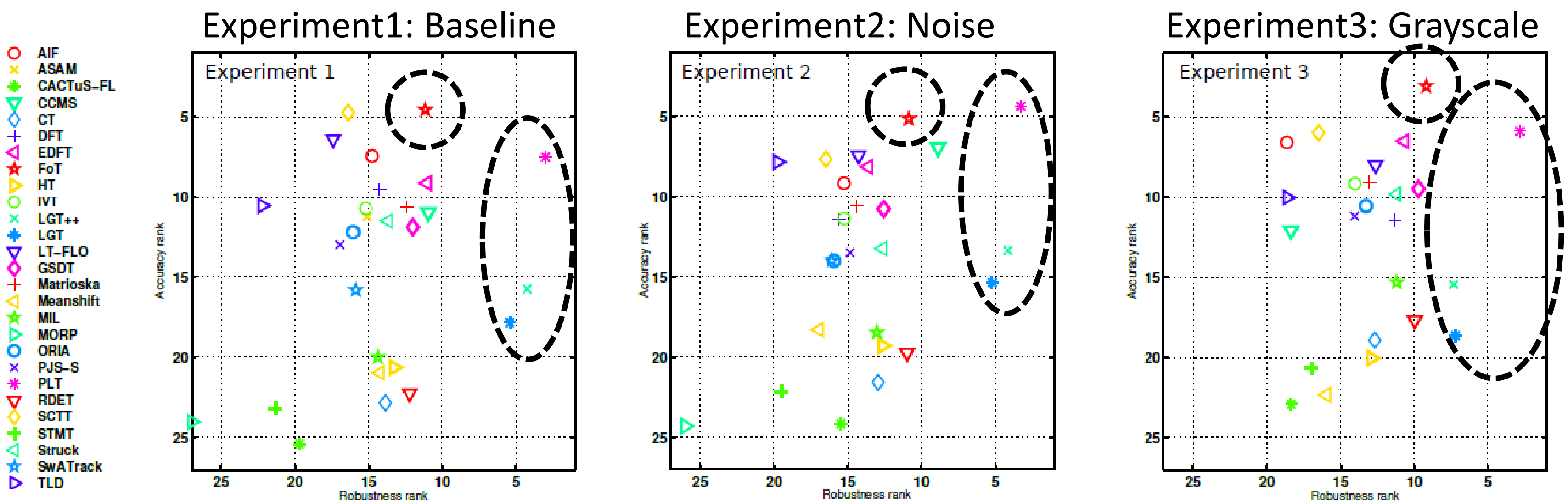
Experiment1: Baseline Experiment2: Noise Experiment3: Grayscale



	R_Σ
PLT*	4.48
FoT*	7.33
EDFT*	9.85
LGT++*	10.05
LT-FLO	11.02
GSDT	11.07
SCTT	11.29
CCMS*	11.36
LGT*	11.61
Matrioska	11.68
AIF	11.98
Struck*	12.01
DFT	12.25
IVT*	12.62
ORIA*	13.66
PJS-S	13.92
TLD*	14.83
MIL*	15.38
RDET	15.48
HT*	16.46
CT*	17.13
Meanshift*	18.12
SwATrack	19.29
STMT	20.63
CACTuS-FL	20.99
ASAM	22.39
MORP	25.89

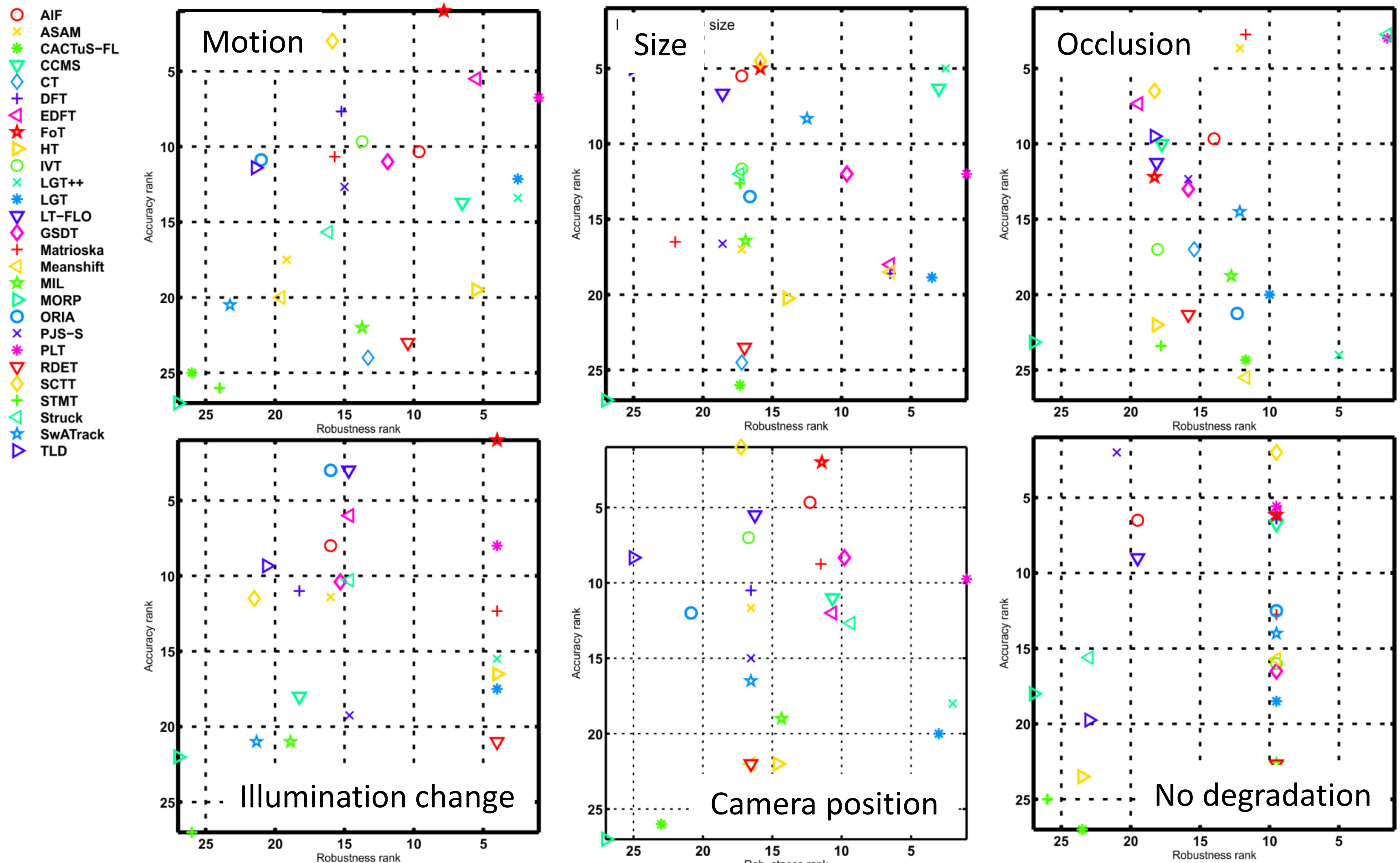
Results: Experiments 1,2,3

- In all experiments PLT best in robustness
- In Baseline and Noise, LGT++ and LGT tightly follow
  - Three trackers perform quite well even in noisy initializations
- But in accuracy, the top performing is FoT except in Noise



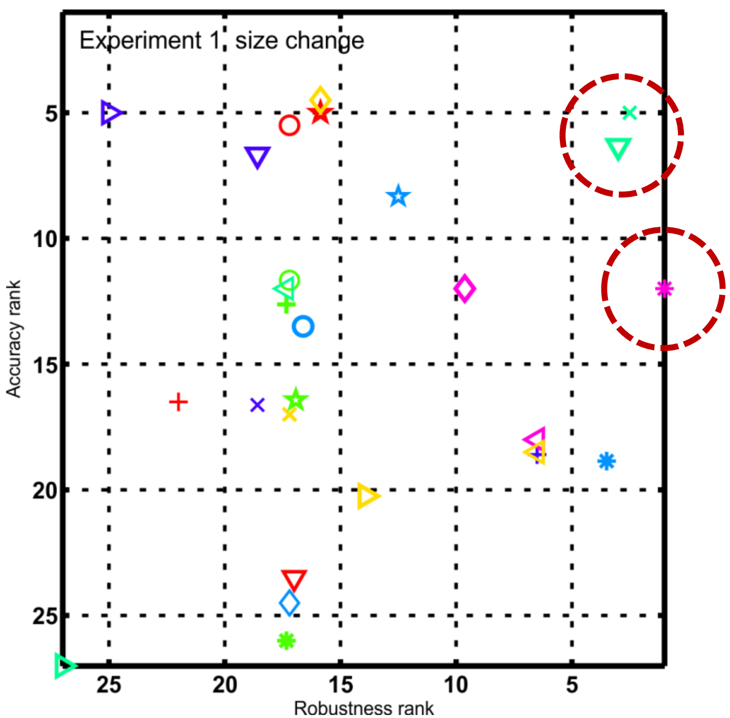
Performance w.r.t. attributes (Ex1)

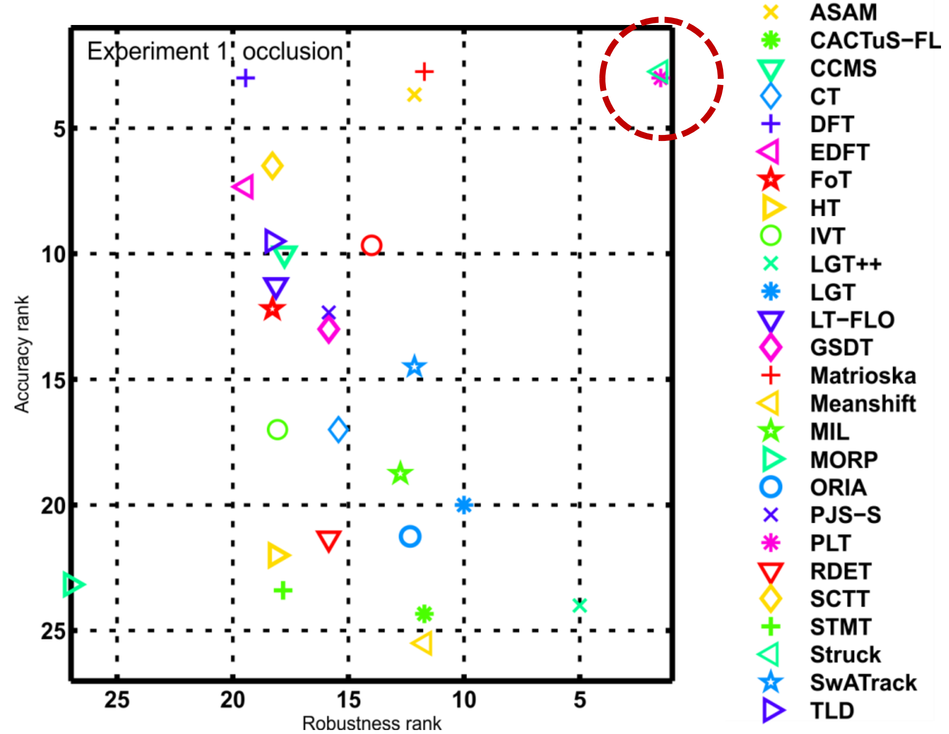
- Average top-performing remain at the top, but...



Performance w.r.t. attributes (Ex1)

- Size change:
  - Best robustness still PLT
  - Best tradeoff between robustness and accuracy: LGT++, CCMS
- Occlusion:
  - PLT and STRUCK best tradeoff





Tracking speed

- Calculated frame rate
- Note! This depends on HW/SW

- PLT (C++) ~169fps
- FoT (C++) ~156fps
- CCMS (Matlab) ~57fps

*Results not verified yet!
Wait for the journal version.

	FPS	Implem.	Hardware
PLT	169.59	C++	Intel Xeon E5-16200
FoT	156.07	C++	Intel i7-3770
EDFT	12.82	Matlab	Intel Xeon X5675
LGT++	5.51	Matlab / C++	Intel i7-960
LT-FLO	4.10	Matlab / C++	Intel i7-2600
GSDT	1.66	Matlab	Intel i7-2600
SCTT	1.40	Matlab	Intel i5-760
CCMS	57.29	Matlab	Intel i7-3770
LGT	2.25	Matlab / C++	AMD Opteron 6238
Matrioska	16.50	C++	Intel i7-920
AIF	30.64	C++	Intel i7-3770
Struck	3.46	C++	Intel Pentium 4
DFT	6.65	Matlab	Intel Xeon X5675
IVT	5.03	Matlab	AMD Opteron 6238
ORIA	1.94	Matlab	Intel Pentium 4
PJS-S	1.18	Matlab / C++	Intel i7-3770K
TLD	10.61	Matlab	Intel Xeon W3503
MIL	4.45	C++	AMD Opteron 6238
RDET	22.50	Matlab	Intel i7-920
HT	4.03	C++	Intel i7-970
CT	9.15	Matlab / C++	Intel Pentium 4
Meanshift	8.76	Matlab	Intel Xeon
SwATrack	2.31	C++	Intel i7
STMT	0.24	C++	Intel Xeon X7460
CACTuS-FL	0.72	Matlab	Intel Xeon X5677
ASAM	0.93	Matlab	Intel i5-2400
MORP	9.88	Matlab	Intel i7

Visual degradation ranking

- Median over accuracy and robustness over all trackers

	camera	illum.	occl.	size	mot.	nondeg
Acc.	0.57	0.57	0.58	0.42	0.57	0.61
Rob.	1.58	0.56	0.66	0.93	0.85	0.00

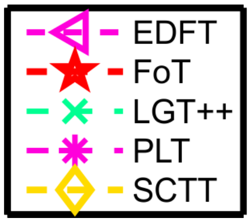
- No degradation simplest (accuracy and robustness)
- Robustness:
  - Camera motion and Object size change seem the most challenging (lots of failing)
- Accuracy:
  - Size change most challenging.
  - Folowed by Camera motion, Illumination, Object motion, and Occlusion.

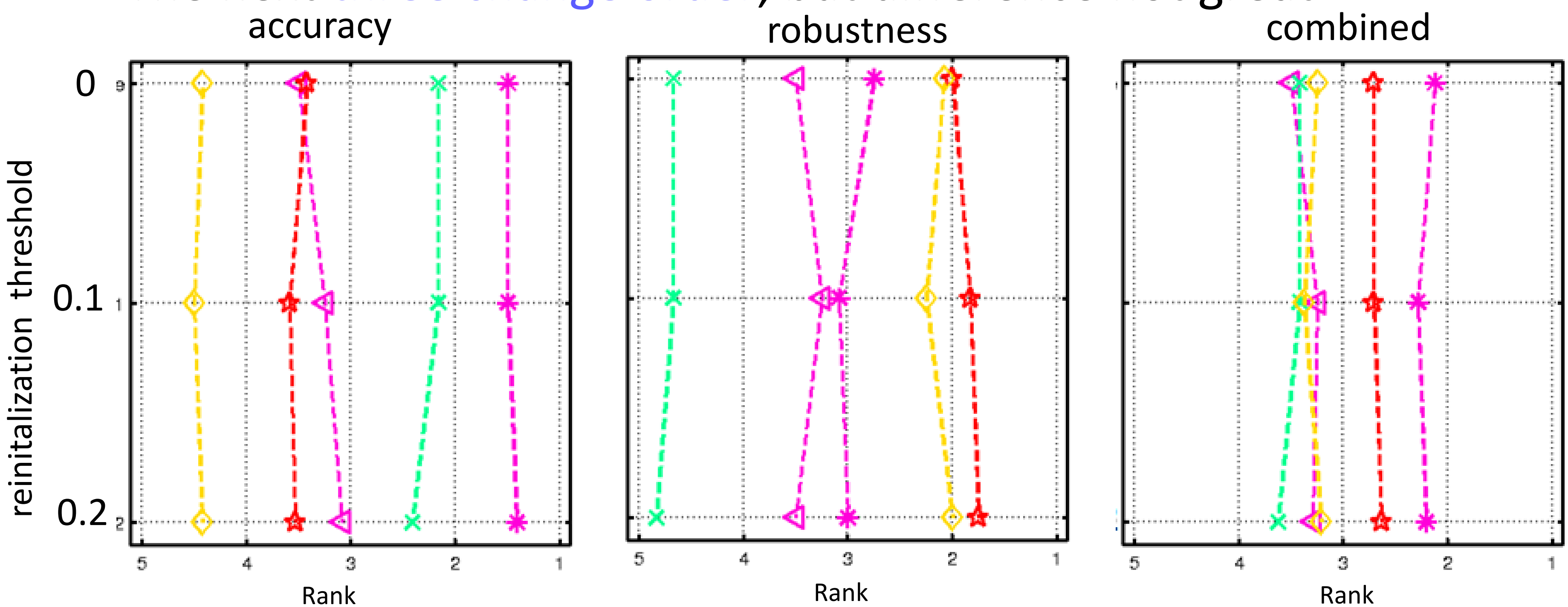
experiments and results

ADDITIONAL VOT2013 EXPERIMENTS

Effects of failure thresholds

- Repeated Experiment 1 with top-performing trackers
- Reinitialization threshold varied (0,0.1,0.2)
- Authors provided the binaries/code of their trackers
- Top two trackers remain at the top
- The next three change order, but difference not great





Additional VOT2013 experiments

- Performed variation of the Experiment 1 with the five top-performing trackers
  - *LT-Flo was excluded from evaluation due to crashing*

1. Dropping frames:
   - Dropping every 3rd frame.
2. Blank frames:
   - Replace each 5th frame with a black frame.
3. Resize:
   - Resize all images to 60%.
4. Reverse:
   - Reverse the order of frames in each sequence.

Additional VOT2013 experiments

- Baseline:
- Reverse:
- Average over all:
- Big shift in ranking: Blank frames
- Largest jump in ranking: EDFT

Baseline:

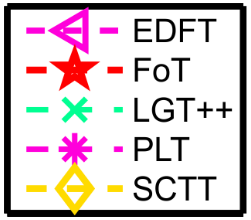
PLT	FoT	SCTT	LGT++	EDFT
2.12	2.71	3.25	3.42	3.5

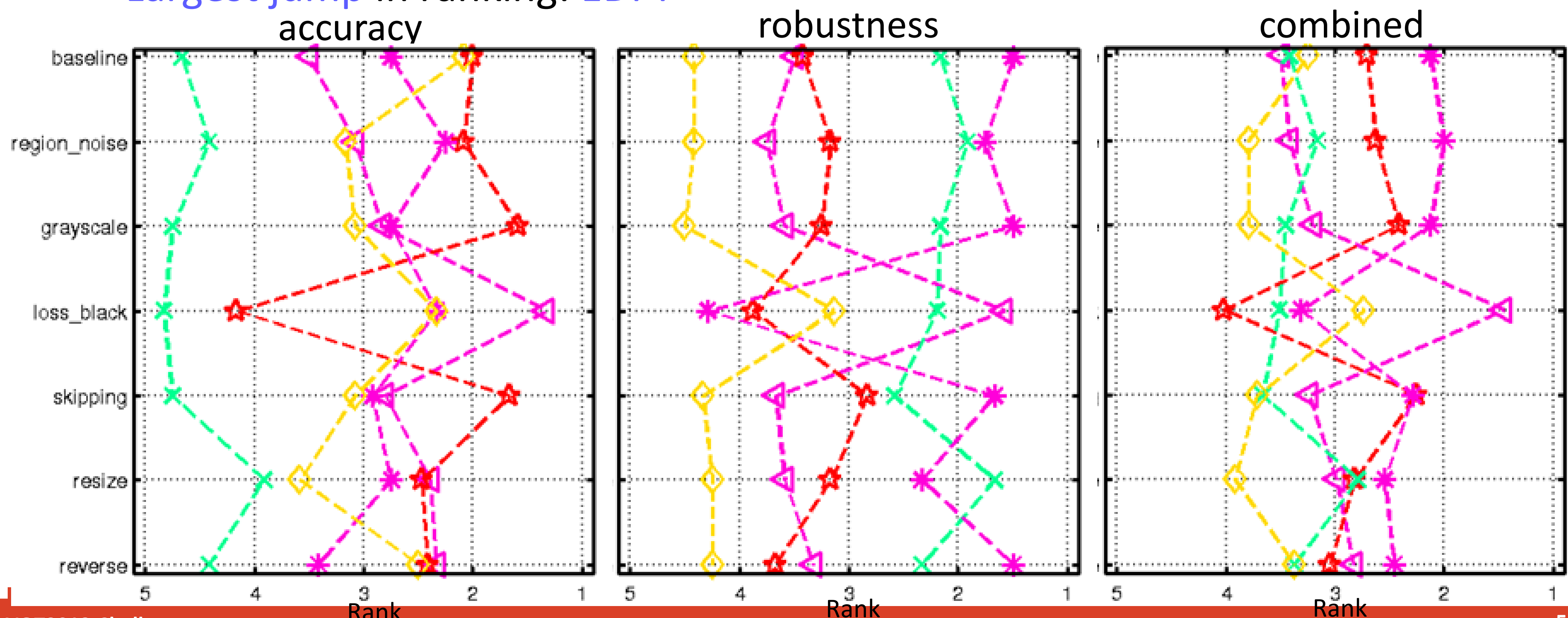
Reverse:

PLT	EDFT	FoT	LGT++	SCTT
2.46	2.83	3.04	3.38	3.38

Average over all:

PLT	FoT	EDFT	LGT++	SCTT
2.38	2.85	2.95	3.35	3.52

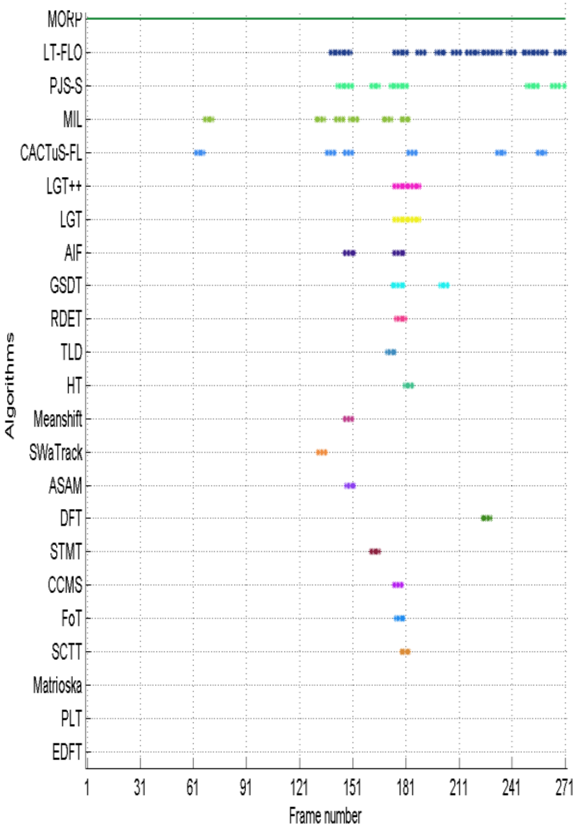


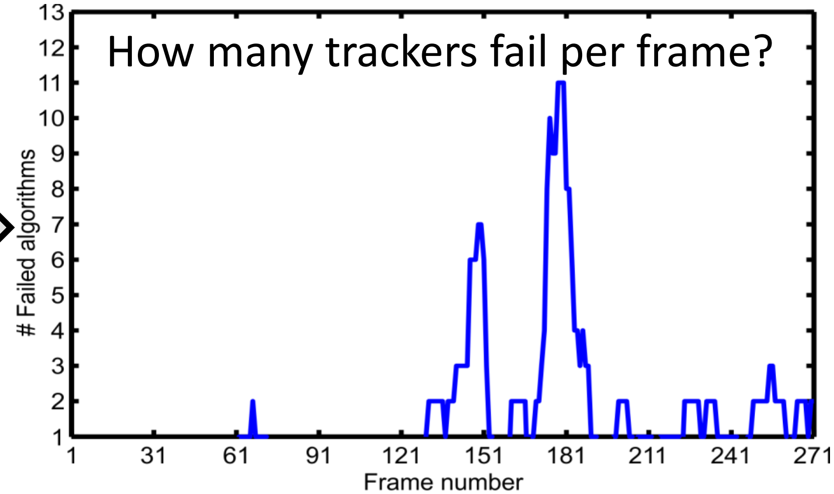


Sequence ranking

- For each sequence calculated how many times each tracker failed at least once in each frame

failure frames for bycicle





Sequence ranking

- Challenging: bolt, hand, diving, gymnastics
- Itermediate: torus, skater
- Surprise: Less challenging David and Singer (overfitting?)
- Easiest: Cup

Sequence	Baseline (Av)	Baseline (Max)	Baseline (Frame)
bolt	4,28	13	242
diving	4,23	9	105
hand	4,22	14	51
gymnastics	3,13	12	98
woman	2,86	15	565
sunshade	2,79	11	85
torus	2,67	8	189
iceskater	2,38	6	227
singer	1,68	4	268
david	1,36	4	337
face	1,22	3	140
bicycle	1,22	11	178
juice	1,12	4	242
jump	0,93	4	203
car	0,92	5	253
cup	0,22	2	232

- Locality: a sequence may be challenging only locally

Sequence ranking: Challenging

bolt
(camera motion, object motion)

hand
(object motion and size change)

diving (most challenging part)
(camera motion at the end, size change)

gymnastic (most challenging part)
(camera and object motion + size change)

Sequence
bolt
diving
hand
gymnastics
woman
sunshade
torus
iceskater
singer
david
face
bicycle
juice
jump
car
cup

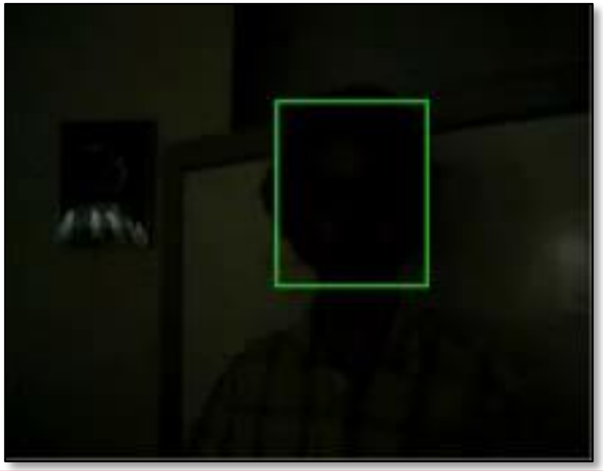
Sequence ranking: Other

- Intermediate (torus, skater)

(object motion)

(camera motion, size change)

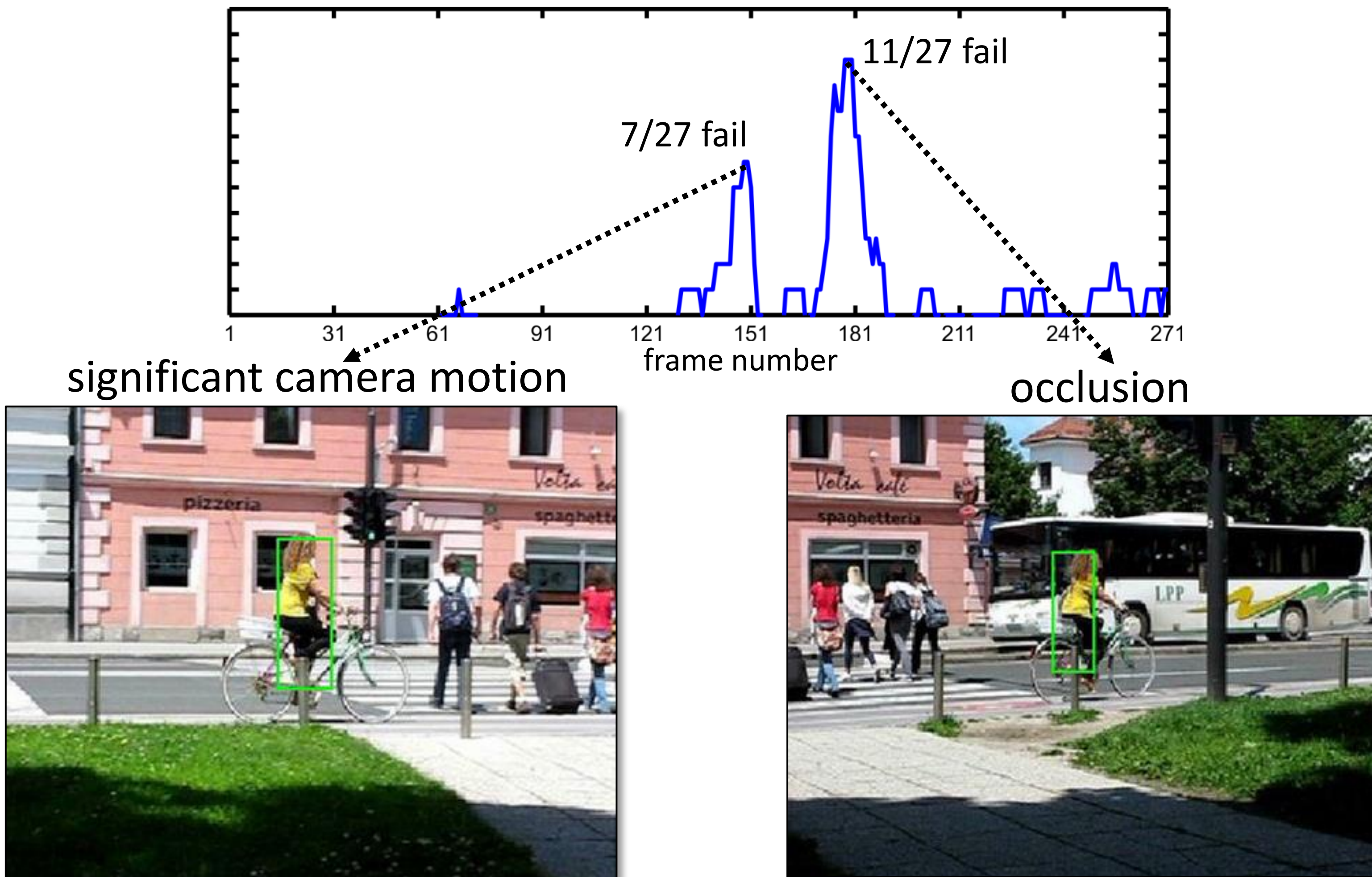
- Less challenging (David and Singer)



Sequence
bolt
diving
hand
gymnastics
woman
sunshade
torus
iceskater
singer
david
face
bicycle
juice
jump
car
cup

Sequence ranking: Locality

- Bicycle: on average not challenging, but very challenging at particular frames where many trackers fail



THE VOT2013 ONLINE RESOURCES

http://votchallenge.net

Summary

- Dataset
  - Considered diversity of visual properties
  - Per-frame annotation of frame attributes
- Evaluation system
  - Multiple platforms
  - Documented tracker integration
- Performance measures
  - Accuracy + Robustness
  - Rank-based comparison methodology
- Analysis of the dataset and the trackers

Summary

- Sparse discriminative PLT quite well in robustness
  - Does not address the size change → accuracy decreases when the object size is significantly changing
- Part-based trackers with rigid constellation
  - Better accuracy at reduced robustness
- Relaxing constellation
  - Increases robustness, but may significantly decrease the accuracy
- Good tradeoffs are still achieved by global visual models, dynamic models may help a great deal.
- Some sequences apparently less challenging
  - Significant camera + object motion + size change challenging
- VOT2013 Challenge winner PLT

Note: we consider sparse trackers as part-based, since they do not apply a global visual model.

Thanks

- The VOT2013 committee*

- Everyone who participated!

et al.: Adam Gatt (DSTO), Ahmad Khajenezhad (Sharif University of Technology), Ahmed Salahledin (Nile University), Ali Soltani-Farani (Sharif University of Technology), Ali Zarezade (Sharif University of Technology), Alfredo Petrosino (Parthenope University of Naples), Anthony Milton (University of South Australia), Behzad Bozorgtabar (University of Canberra), Bo Li (Panasonic R&D Center), Chee Seng Chan (University of Malaya), CherKeng Heng (Panasonic R&D Center), Dale Ward (University of South Australia), David Kearney (University of South Australia), Dorothy Monekosso (University of West England), Hakki Can Karaimer (Izmir Institute of Technology), Hamid R. Rabiee (Sharif University of Technology), Jianke Zhu (Zhejiang University), Jin Gao (National CAS), Jingjing Xiao (University of Birmingham), Junge Zhang (Chinese Academy of Sciences), Junliang Xing (CAS), Kaiqi Huang (Chinese Academy of Sciences), Karel Lebeda (University of Surrey), Simon Hadfield (University of Surrey), Lijun Cao (Chinese Academy of Sciences), Mario Edoardo Maresca (Parthenope University of Naples), Mei Kuan Lim (University of Malaya), Mohamed ELHelw (Nile University), Michael Felsberg (Linkoeping University), Paolo Remagnino (Kingston University), Richard Bowden (University of Surrey), Roland Goecke (Australian National University), Rustam Stolkin (University of Birmingham), Samantha YueYing Lim (Panasonic R&D Center), Sara Maher (Nile University), Sebastien Poullot (NII), Sebastien Wong (DSTO), Shin ichi Satoh (NII), Weihua Chen (Chinese Academy of Sciences), Weiming Hu (CAS), Xiaoqin Zhang (CAS), Yang Li (Zhejiang University), ZhiHeng Niu (Panasonic R&D Center)

This work was supported in part by the Slovenian research agency programs and projects P2-0214, P2-0094, J2-4284, J2-3607, J2-2221 and EU-project 257906, CTU Project SGS13/142/OHK3/2T/13 and the Technology Agency of the Czech Republic project TE01020415 V3C – Visual Com.

*sorted by authors order of this presentation