VOT2014
workshop on visual object tracking challenge
September, 6th 2014 - Zurich, Switzerland

The Visual Object Tracking VOT2014: Challenge and results

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

University of Ljubljana
Faculty of Computer and Information Science

AIT Austrian Institute of Technology

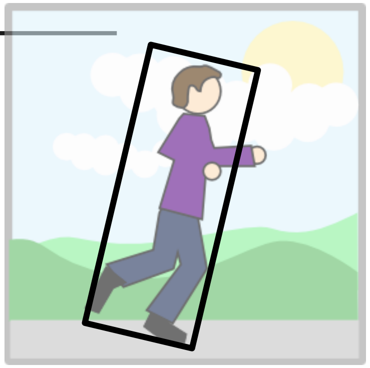
University of Birmingham

Outline

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

SCOPE OF THE VOT2014 CHALLENGE

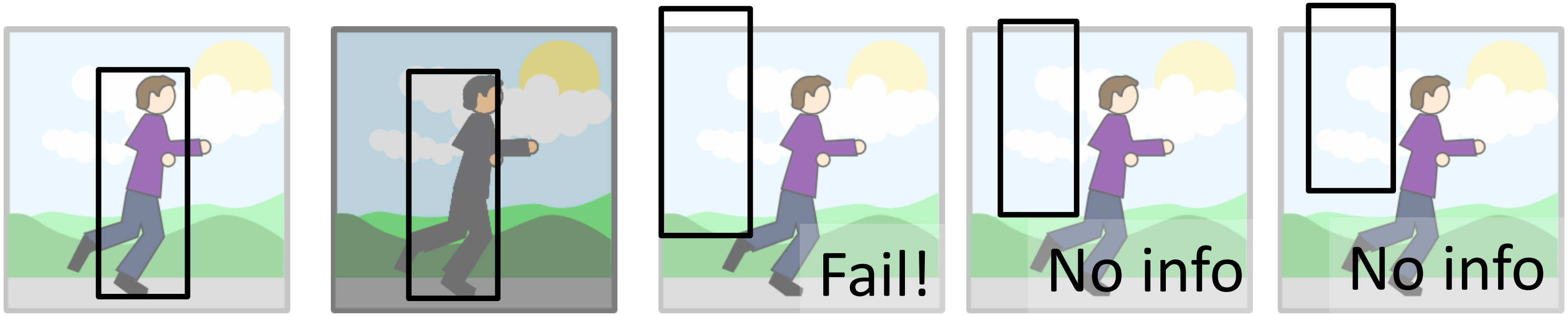
Selected class of trackers



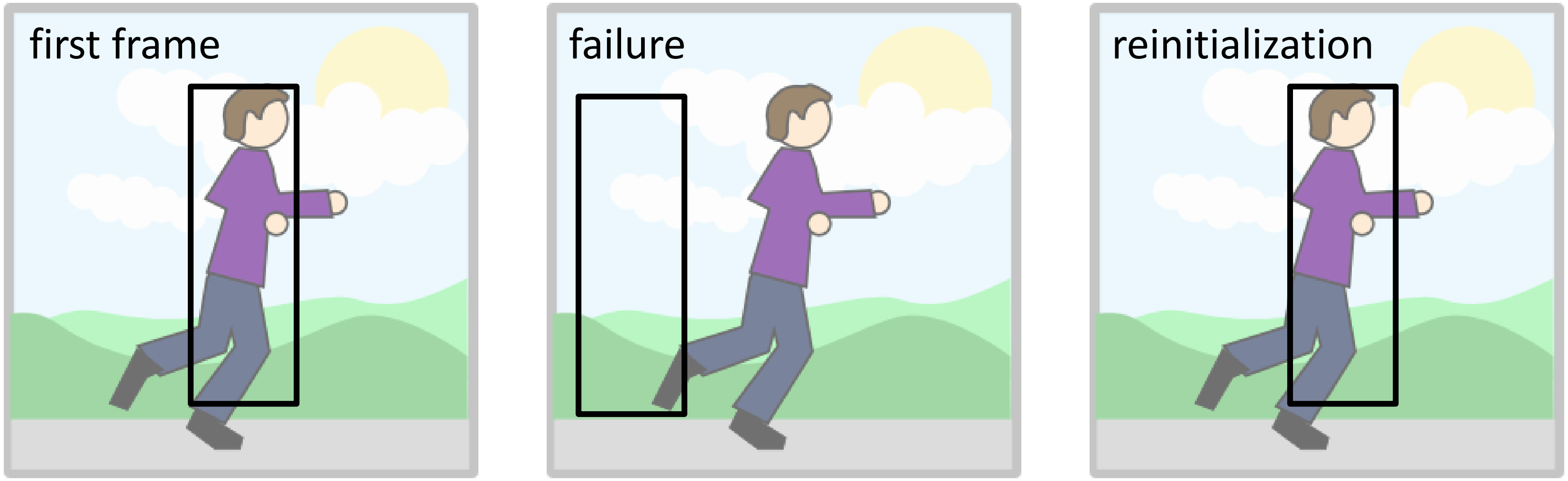
- Single-object, single-camera, model-free, short-term, causal trackers
- Model-free:
  - Nothing but a single training example is provided by the BBox in the first frame
- Short-term:
  - Tracker does not perform re-detection
  - Once it drifts off the target we consider that a failure
- Causality:
  - Tracker does not use any future frames for pose estimation
- Object state defined as rotated bounding box

Requirements for tracker implementation

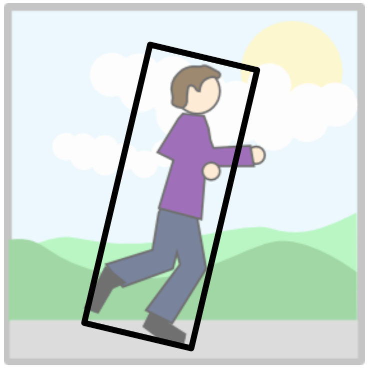
- Would like to use the data fully



- Renitialize once the tracker drifts from the object



Requirements for tracker implementation



- Complete reset:
  - 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
- BB is arbitrary (rotated) rectangle *(new this year!)*
- Parameters may be set internally, but not by detecting a specific sequence
  - Verified for the top-performing trackers
- Submitting a tracker with different parameters not considered a different tracker

Related work on tracker benchmarking

VTES [Smeulders et al., TPAMI2014], *OTB* [Wu, Lim and Yang, CVPR2013],…

- Initialize at beginning, let run till the end. Then compute a performance measure.
- Apply long-term measures (recall, precision) on (mostly) short-term sequences.
- Use of brittle performance measures (like center-based)
- Visual properties of sequences not detailed enough for deep analysis.
- Tracker equivalence not a core problem:
  "If the average overlap for T1 is 0.6 and T2 is 0.61, can we say that T2 is better than T1?"

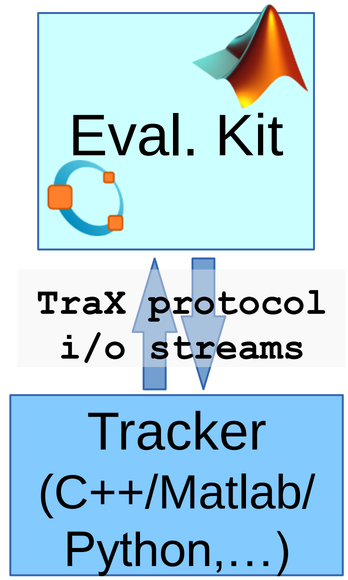
Related work on tracker benchmarking

- VOT2013: 1st short-term tracking challenge
  - Provided fully annotated dataset and evaluation kit
  - Advanced performance evaluation methodology
  - Compared 27 trackers on 16 sequences

- VOT2014 improves on VOT2013 in several aspects:
  - A faster evaluation system
  - Extended dataset
  - Improved performance analysis methodology
  - A system for interactive exploration of results

VOT2014 EVALUATION SYSTEM

VOT2014 Challenge evaluation kit



- Matlab-based kit to automatically perform a battery of standard experiments
- Download from our homepage
  https://github.com/vicoslab/vot-toolkit
- Plug and play!
  - Supports multiple platforms
  - Supports a large variety of programming languages (C/C++/Matlab/Python, etc.)
- Easy to evaluate your tracker on our benchmarks
- Deep integration with tracker→ Fast execution of experiments
- Backward compatibility with VOT2013

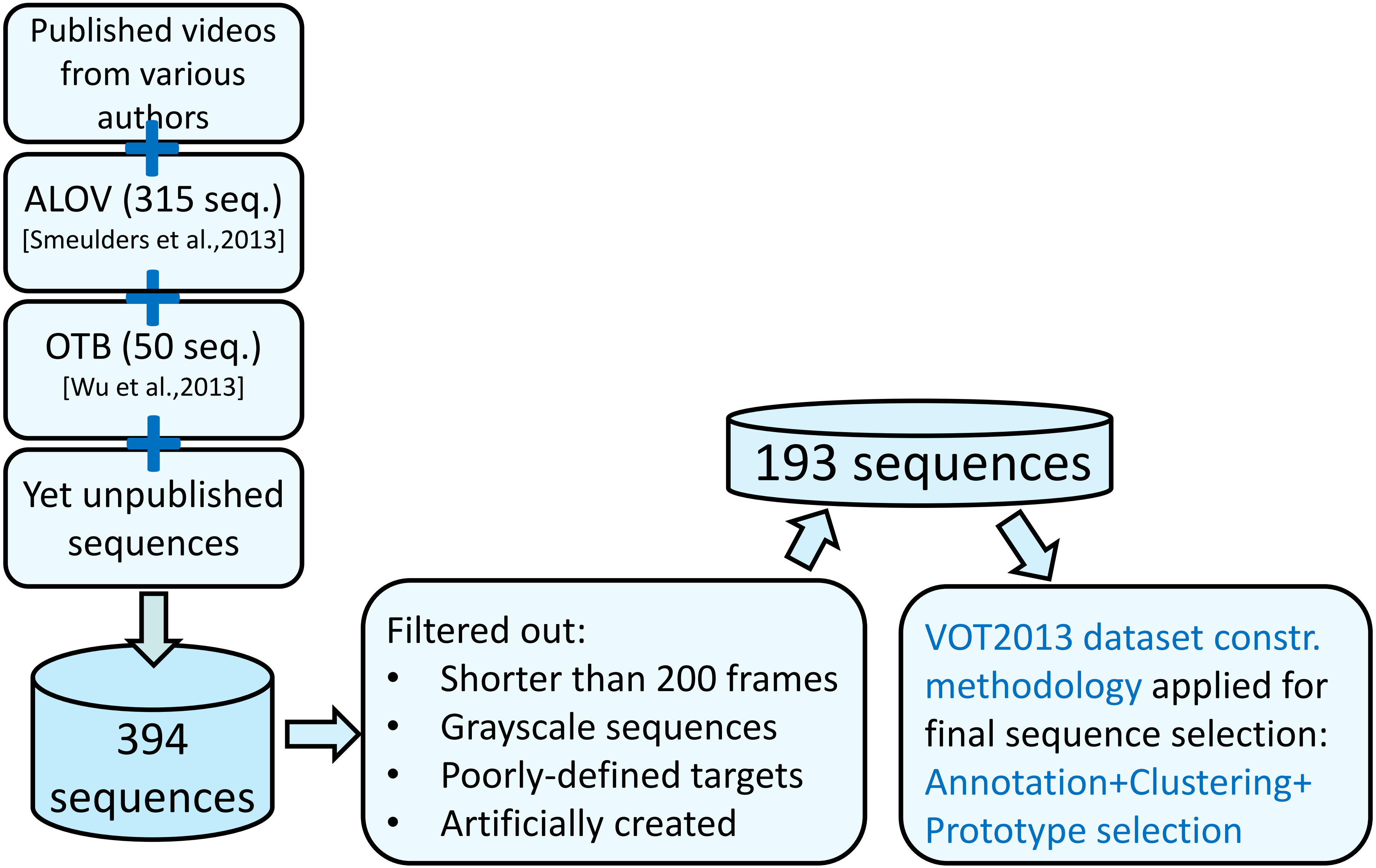
VOT2014 DATASET

Relevant datasets

- Lots of datasets: PETS [Young and Ferryman 2005], CAVIAR[1] , i-LIDS[2], ETISEO[3], CVBASE[4], FERET [Phillips et al., 2000], OTB [Wu et al., 2013], ALOV300+ [Smeulders et al., 2013]
- VOT2013 dataset [Kristan et al., 2013]
  - Contains 16 fully labelled color sequences
  - Diversity in visual attributes
  - Methodology for dataset construction (details in [Kristan et al., 2013])

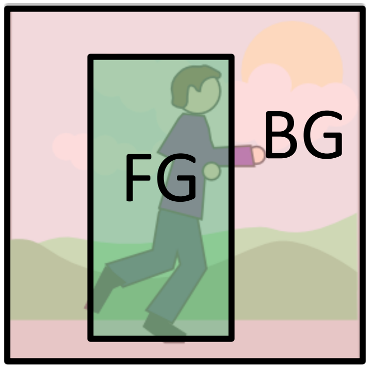
[1] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1
[2] http://www.homeoffice.gov.uk/science-research/hosdb/i-lids
[3] http://www-sop.inria.fr/orion/ETISEO
[4] http://vision.fe.uni-lj.si/cvbase06/

VOT2014 dataset: collection and filtering



VOT2014 dataset: Annotation

- 10 global attributes estimated automatically for 193 sequences
  - Estimators based on ad hoc heuristics
  - Each sequence represented as 10dim feature vector.

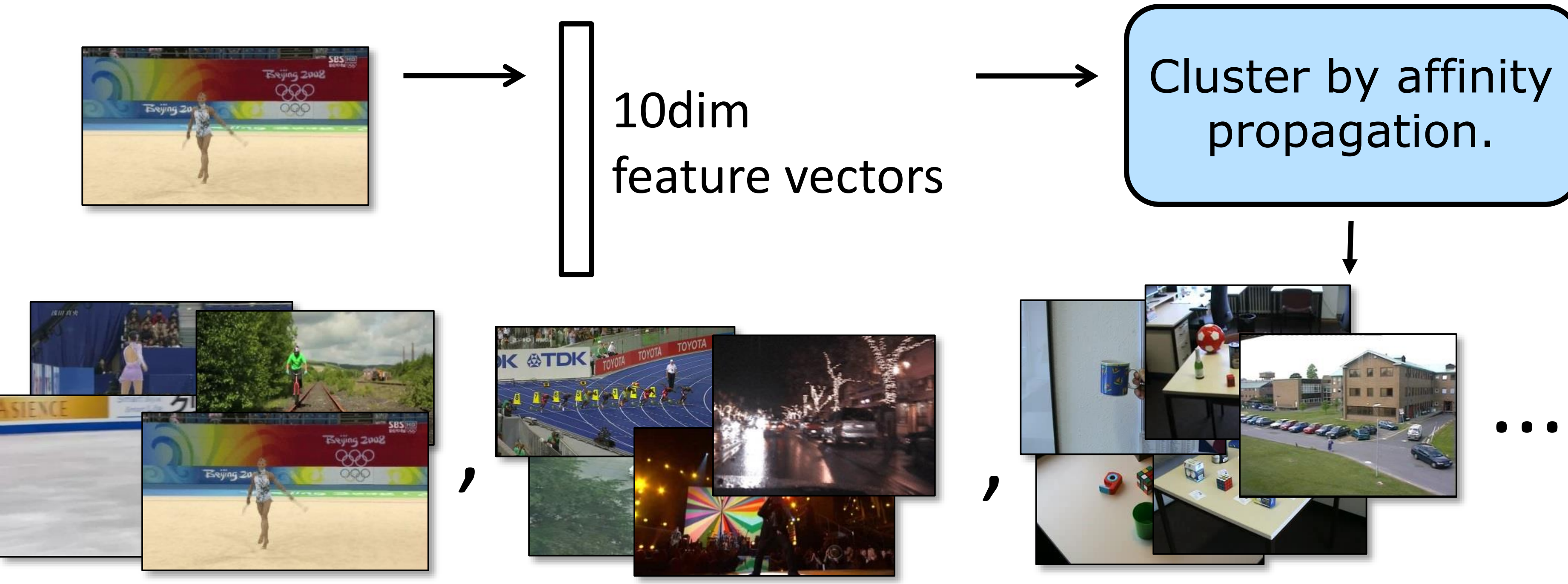


Global attributes:

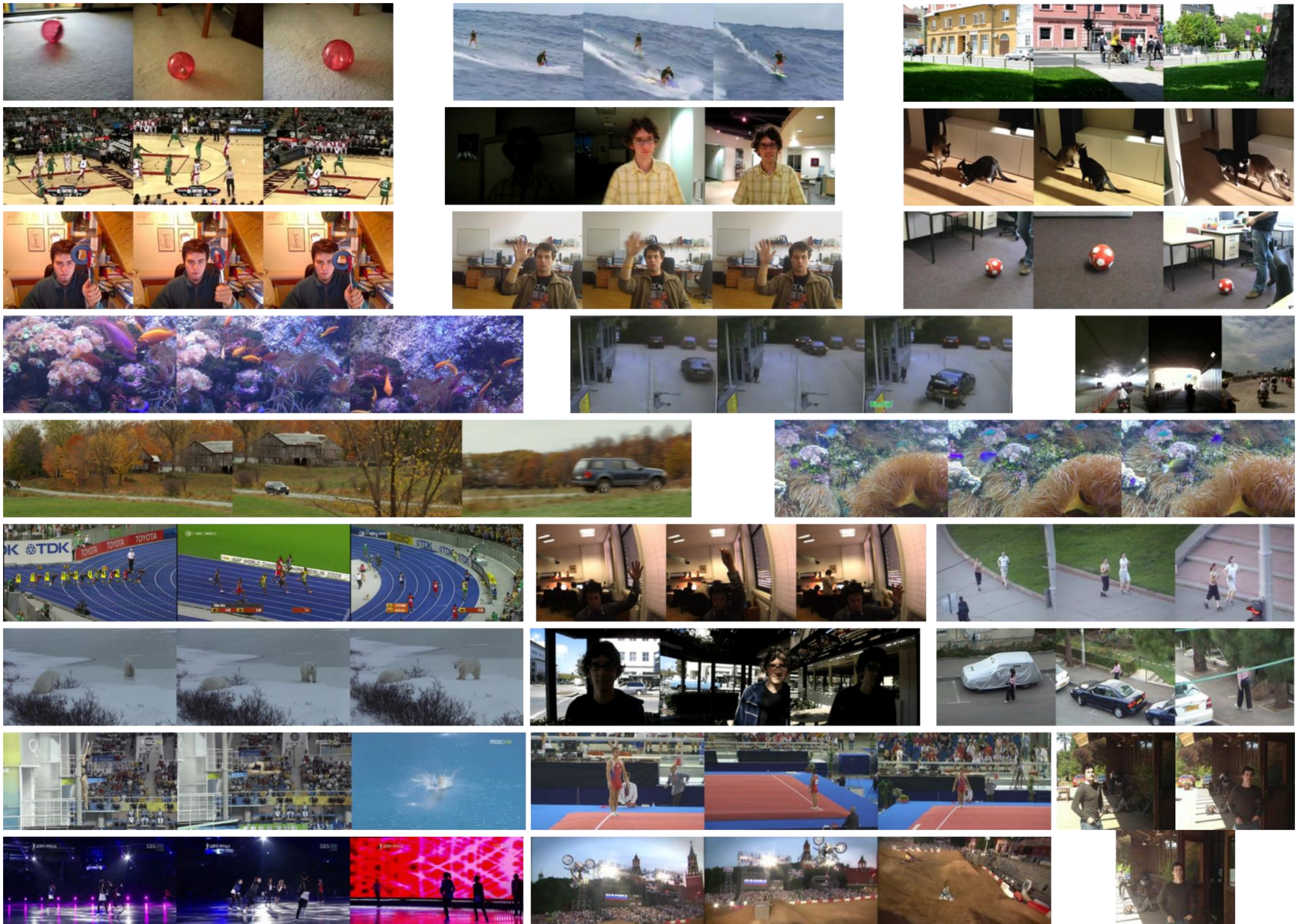
1. Illumination change
   (difference of min/max FG intensity)
2. Size change
   (average of sequential BB size difference)
3. Motion
   (average of sequential BB center difference)
4. Clutter
   (FG/BG color histogram difference)
5. Camera motion
   (patch features motion between frames)
6. Blur
   (Camera focus measure [Kristan et al., 2006])
7. Aspect-ratio change
   (relative to initial BB aspect ratio change)
8. Object color change
   (average hue change inside BB w.r.t initial)
9. Deformation
   (mean intensity change in BB subregions)
10. Scene complexity
    (entropy of grayscale image)

VOT2014 dataset: Clustering

- Sequences clustered into 12 clusters by attributes using Affinity propagation [Frey and Dueck 2007].
- Approx. 2 videos selected from each cluster manually.
  - Make sure that phenomena like occlusion were still well represented.

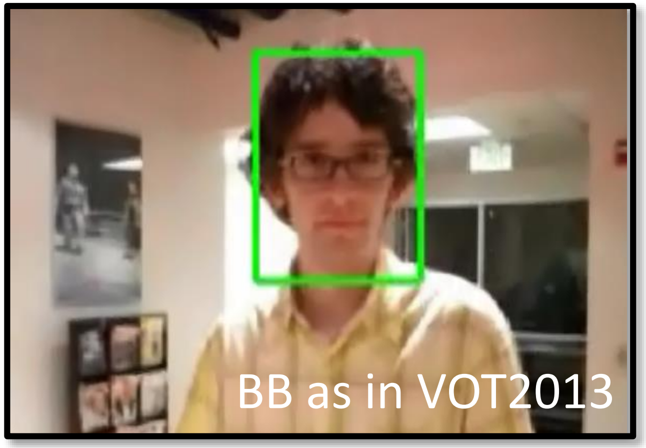


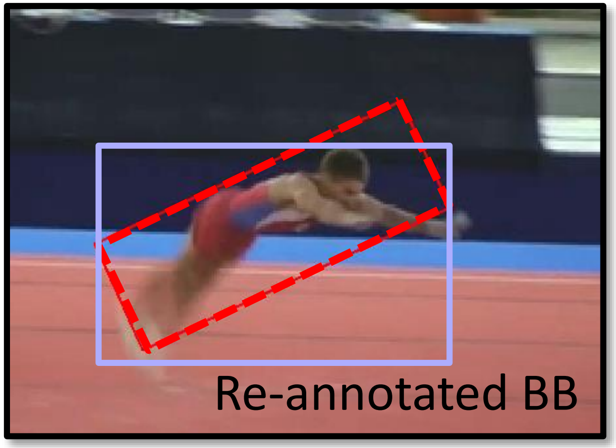
VOT2014 dataset: 25 sequences



VOT2014 dataset – object annotation

- Most came with existing annotation (axis-aligned BB)
- In VOT2013 some were re-annotated (axis-aligned BB)
- In VOT2014, sequences containing elongated, rotating, deforming, objects, re-annotated by rotated BB.
- VOT2014 sequences contain:
  Original, VOT2013 and new rotated BB annotations.





Dataset – frame-level attribute annotation

- Common practice: Each sequence annotated by a visual attribute [Dung et al 2010,Wu et al. 2013]
- However, a visual phenomenon does not last over the entire sequence

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

- For a detailed analysis per-frame annotations required.

VOT2014 dataset – frame annotation

- Manually and semi-automatically 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. Neutral (A)

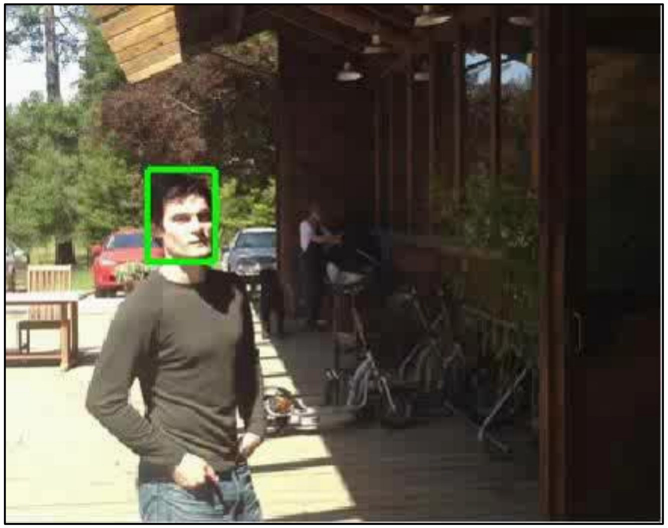
M ... manual annotation, A ... automatic annotation

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

VOT2014 dataset – frame annotation

- Example: Occlusion
  All annotations: occlusion, object size change, camera motion, motion change

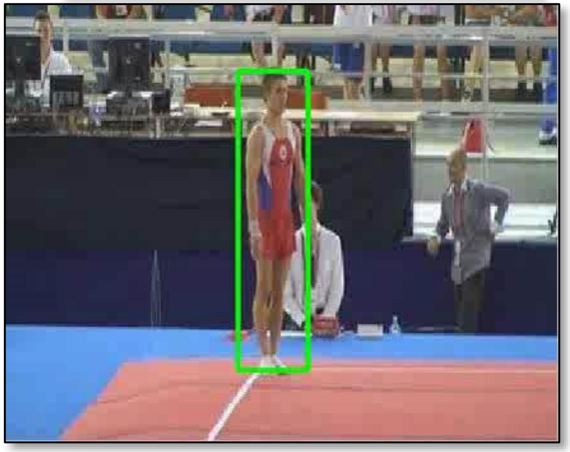
- Example: Illumination change
  All annotations: camera motion, illumination change, motion change



VOT2014 dataset – frame annotation

- Example: Object motion
  Attributes appearing in sequence:
  Motion change, size change

- Example: Camera motion
  Attributes appearing in sequence:
  Camera motion, object motion, size change

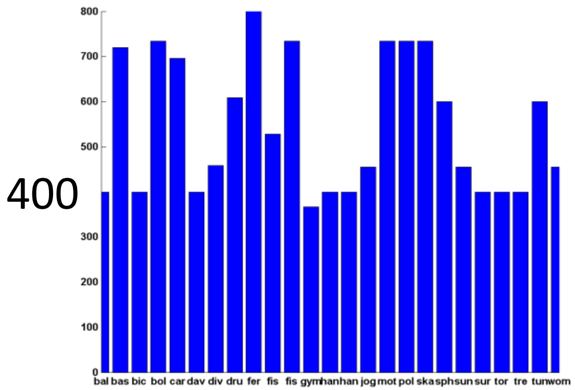


- Example: Size change
  Attributes appearing in sequence:
  Camera motion, illumination change,
  motion change, size change

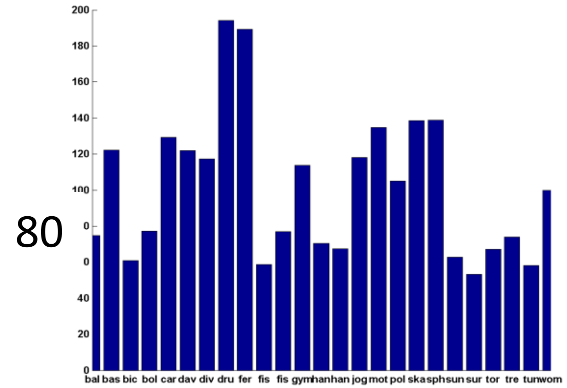
VOT2014 dataset – general stats

- 25 color sequences:

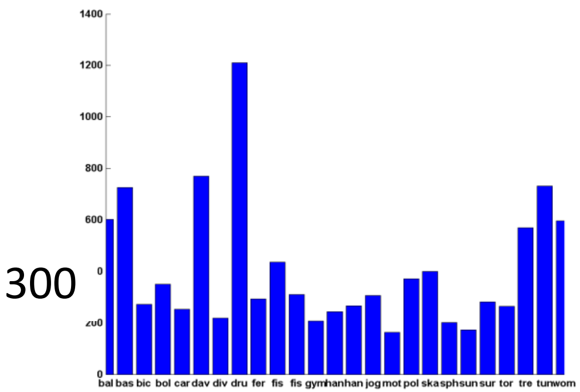
Diagonals of images



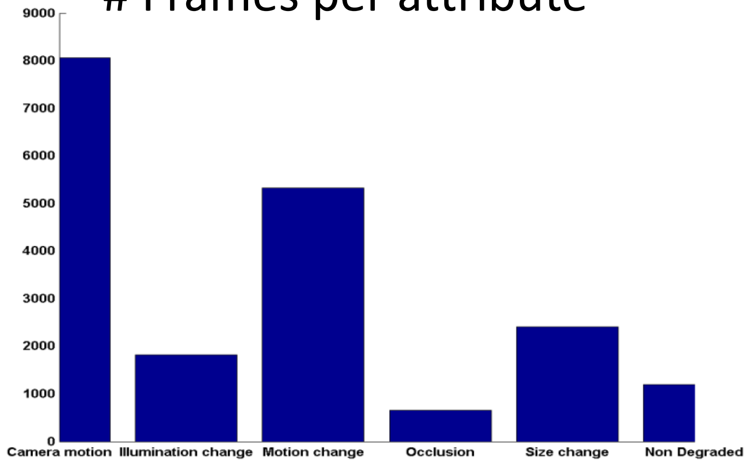
Object bounding box diagonals



Sequence length distribution



Frames per attribute



EVALUATION METHODOLOGY

Performance measures

- Target localization properties measured using the VOT2013 methodology.

- Approach in VOT2013:
  - Interpretability of performance measures
  - Select as few as possible to provide clear comparison

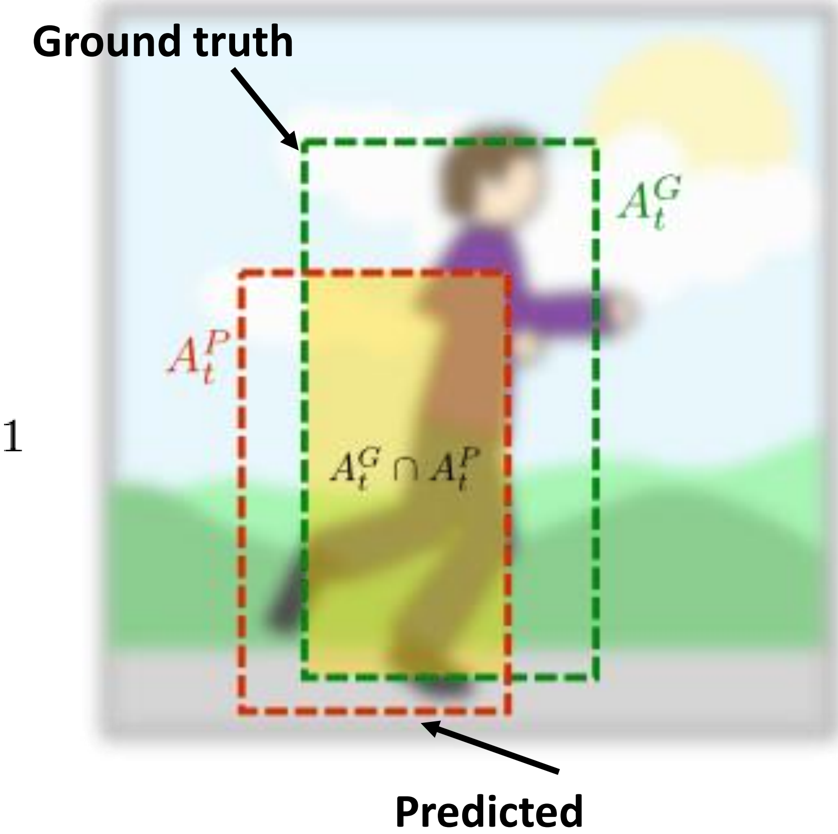
- Based on a recent study[1] two basic weakly-correlated measures are chosen:
  - Accuracy
  - Robustness

[1] Čehovin, Kristan, and Leonardis, "Is my new tracker really better than yours?", WACV2014

VOT2014 measures: Accuracy

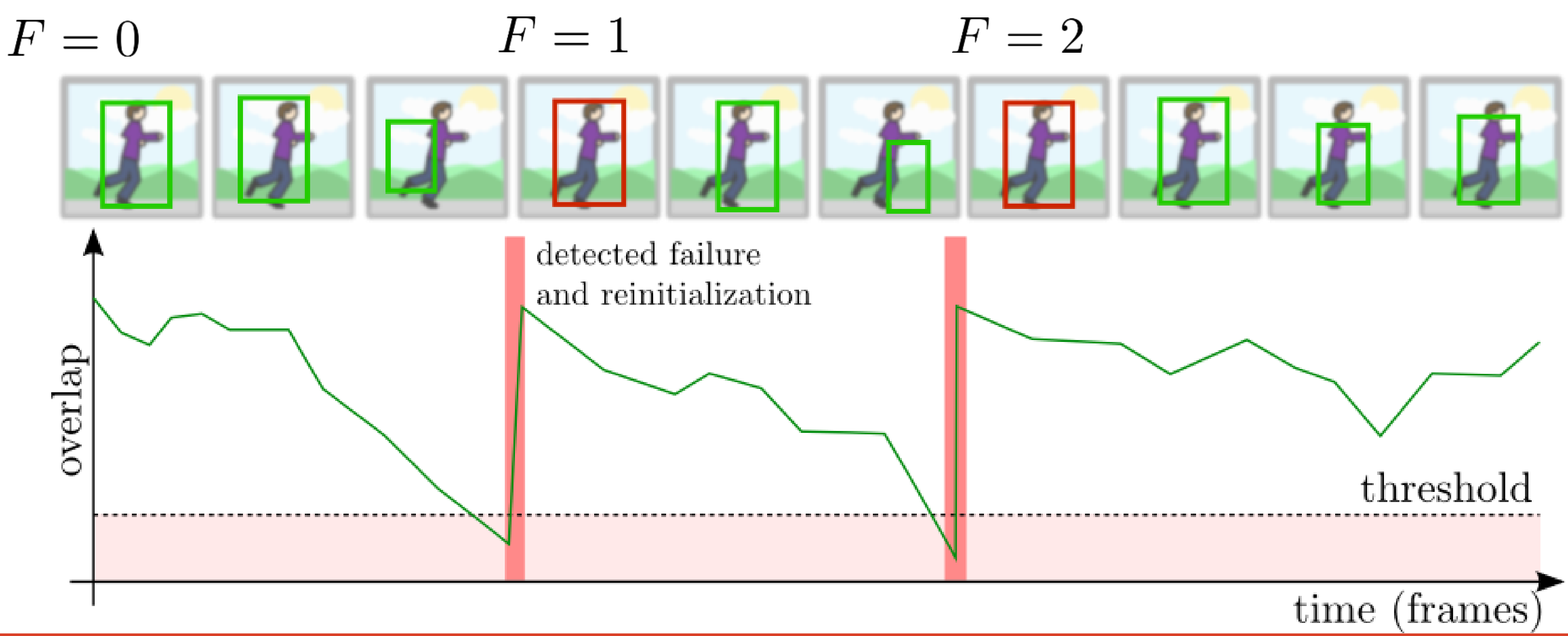
- Overlap between the ground-truth BB and the BB, predicted by a tracker

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



VOT2014 measures: Robustness

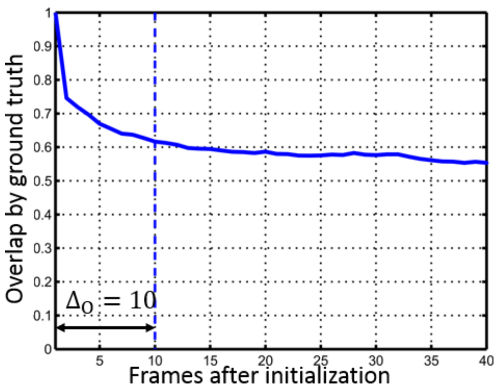
- Counts the number of times the tracker failed and had to be reinitialized
- Failure detected when the overlap $\Phi(\Lambda_G, \Lambda_P)$ drops below a threshold



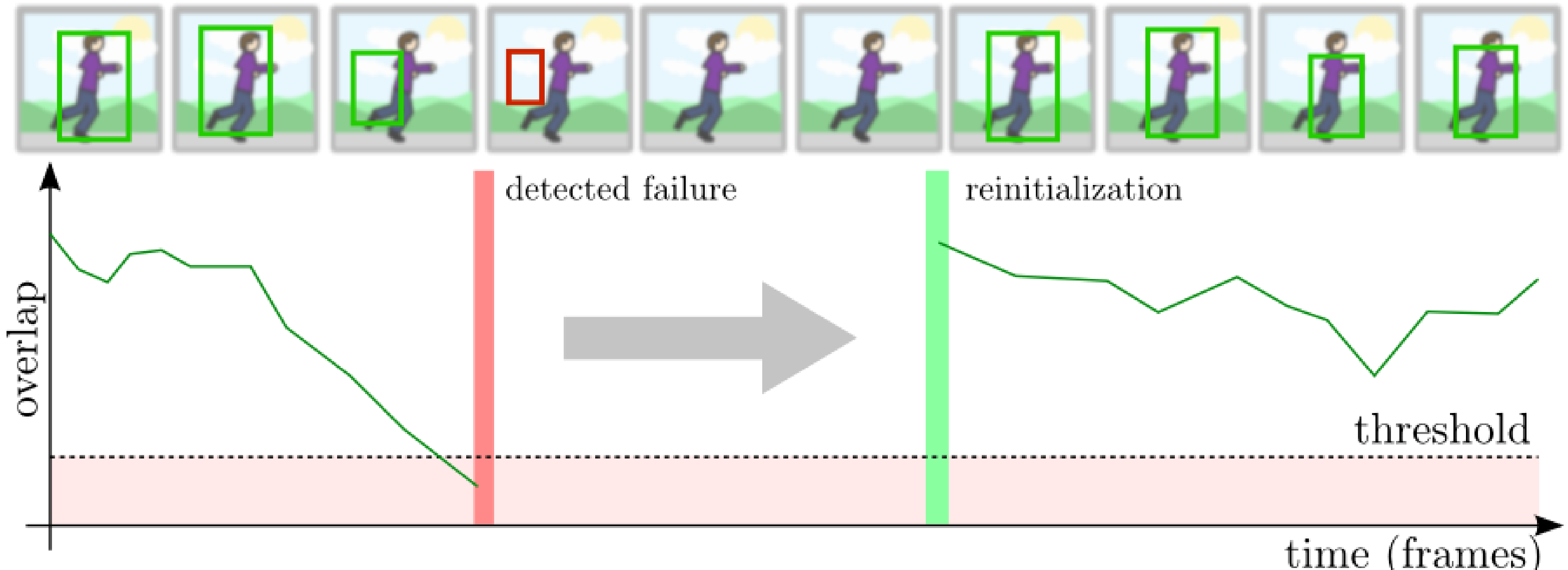
VOT2014 measures: Reinitialization

- 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.



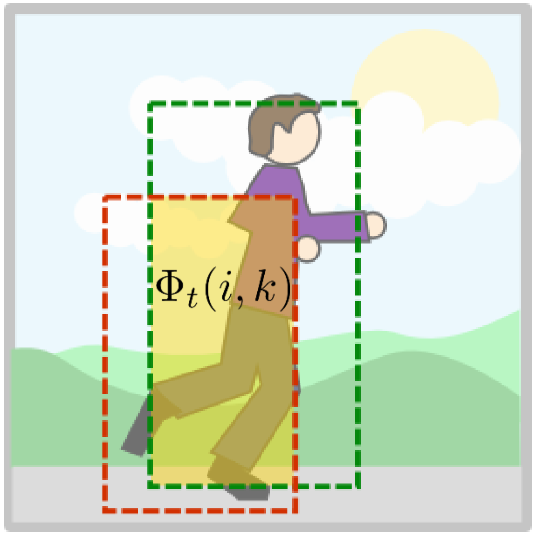
- $\Delta_{\mathrm{O}} = 10$ frames after initialization ignored in accuracy computation to reduce bias in accuracy.



VOT2014 measures: Multiple runs

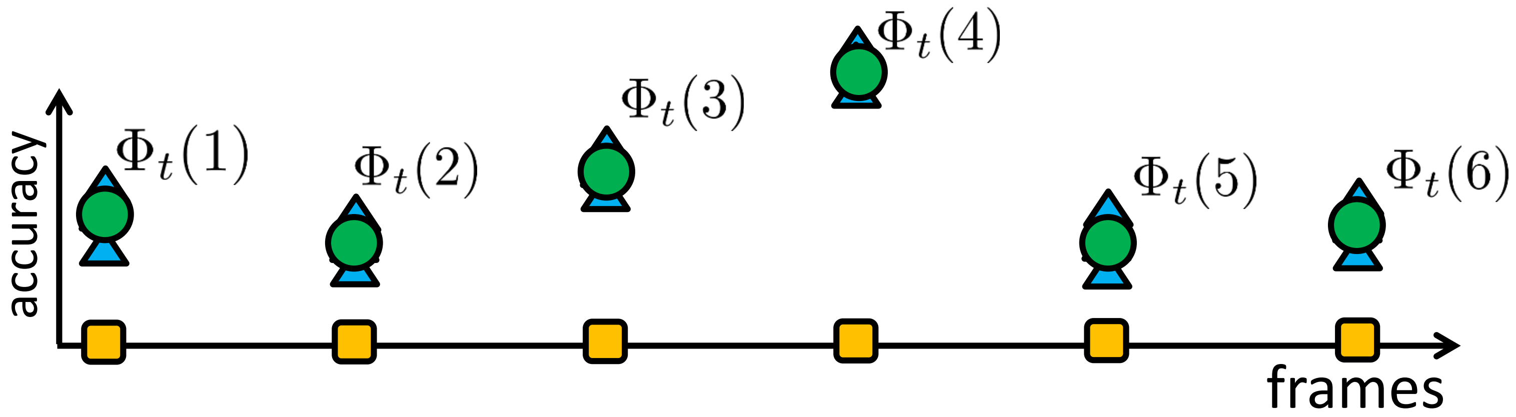
- Measures averaged over multiple runs

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



- Per-frame averaged accuracy

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



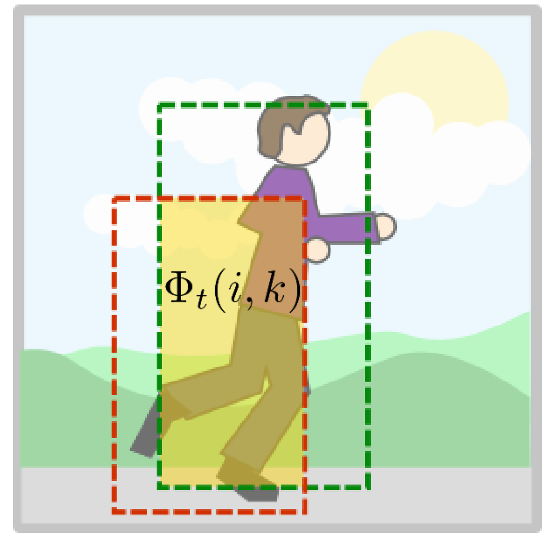
VOT2014 measures: Multiple runs

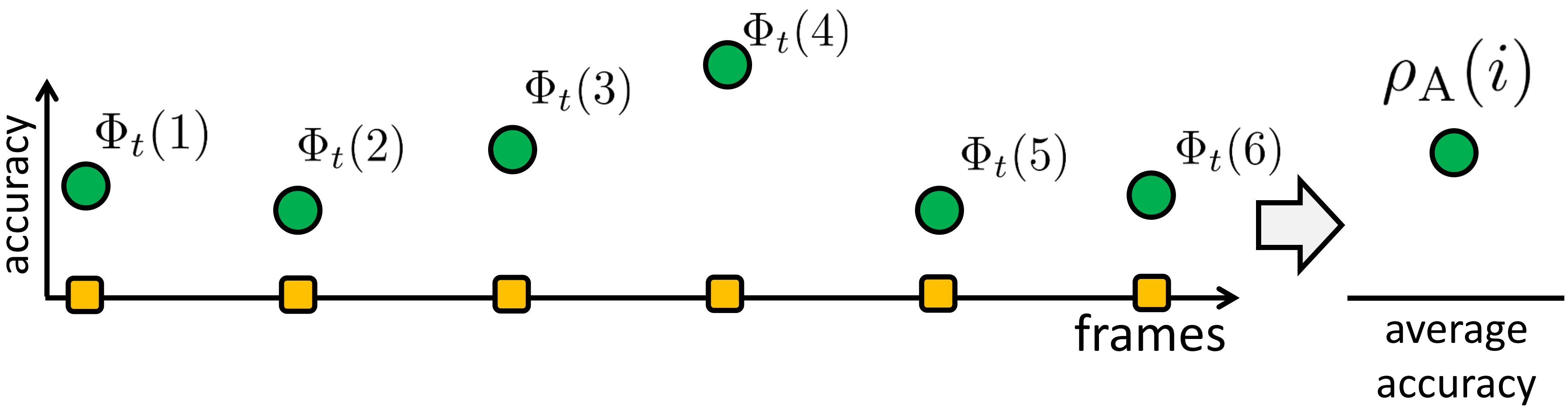
- Average accuracy at frame *t*

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

- Average accuracy over sequence

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

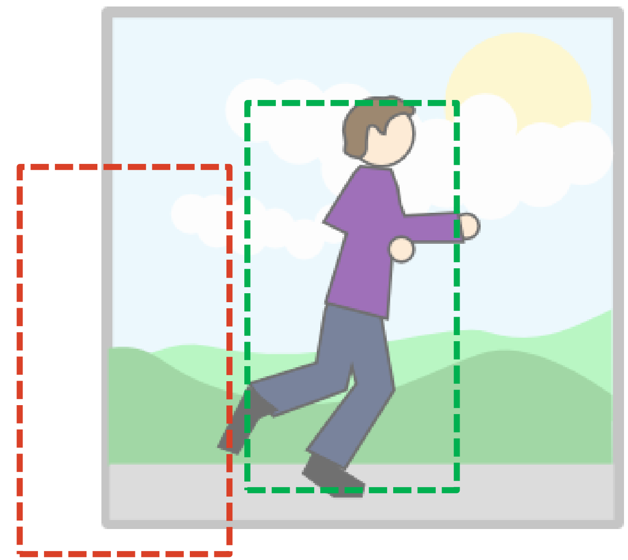




VOT2014 measures: Multiple runs

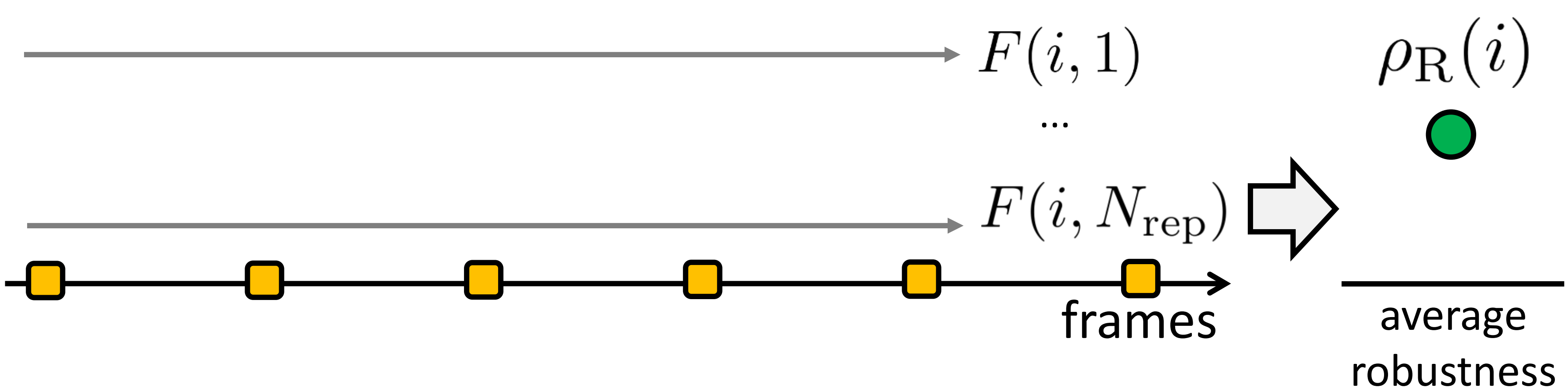
- Multiple measurements of robustness (#failures)

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



- Average robustness per sequence

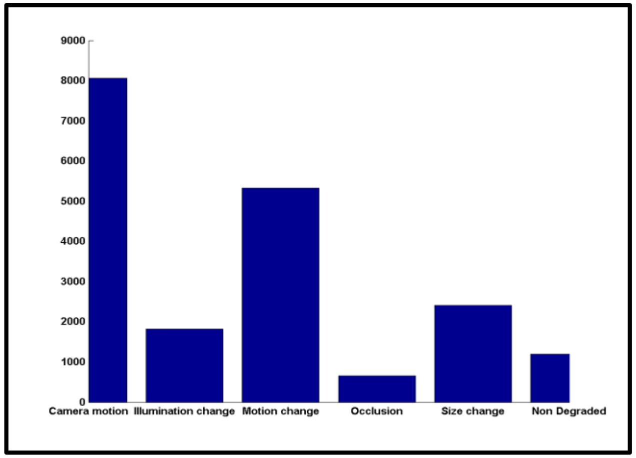
$$\rho_R(i) = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} F(i,k)$$



VOT2014 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 accuracy and robustness 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 performance measure we average the two corresponding rankings

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

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

- 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

- Tests of equality separately for accuracy and robustness

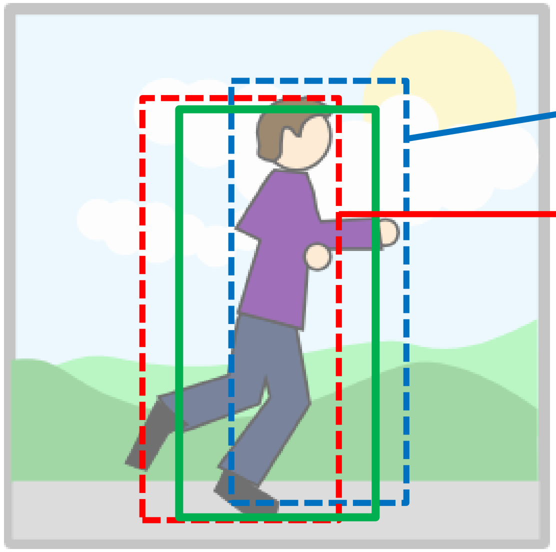
Statistical tests of differences

- VOT2013 introduced tests of statistical significance of differences in tracking performance. (Detalis in [Kristan, 2013])

- Robustness
  - A single robustness measurement per experiment repetition
  - Apply unpaired Wilcoxon Rank-Sum test (Mann-Whitney U-test)

- Accuracy
  - Per-frame measure available for each tracker.
  - Paired Wilcoxon signed-rank test as in [Demšar IJMLR2006]

Accuracy: Practical equivalence



$$\phi_t(i) \gtreqless \phi_t(j)$$

Ground truth ambiguity:

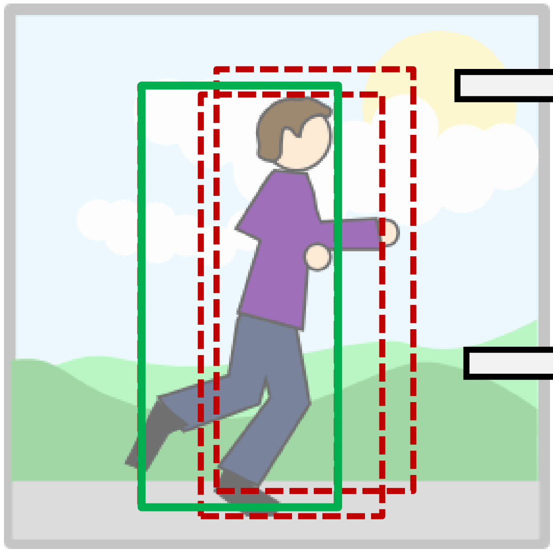
- Noise in annotation
- Multiple ground truth annotations equally valid

- Practical difference:
  "Level of difference that is considered negligibly small"

- A pair of trackers is considered to perform equally well in accuracy if it fails either (1) statistical difference test or (2) practical difference test.

Estimation of practical difference thresholds

- Consider per-frame estimation:



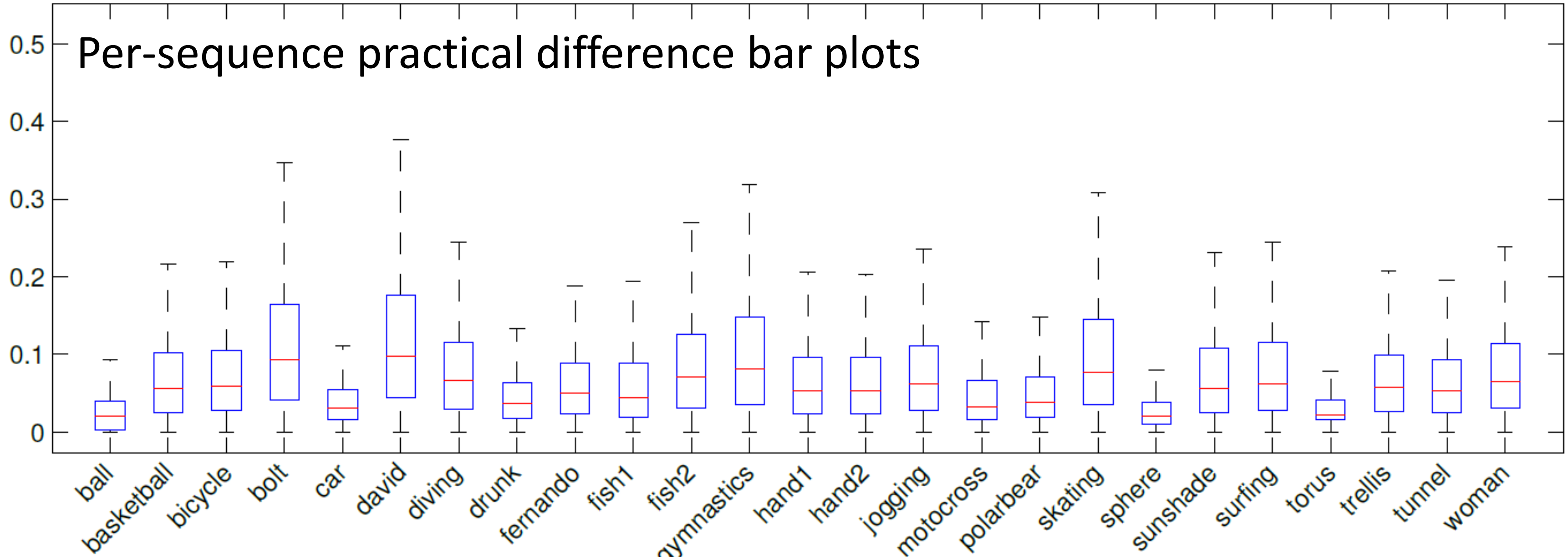
Consider one bounding box a GT and compute N-1 overlaps.

Cycling over all bounding boxes increases then number of overlaps to

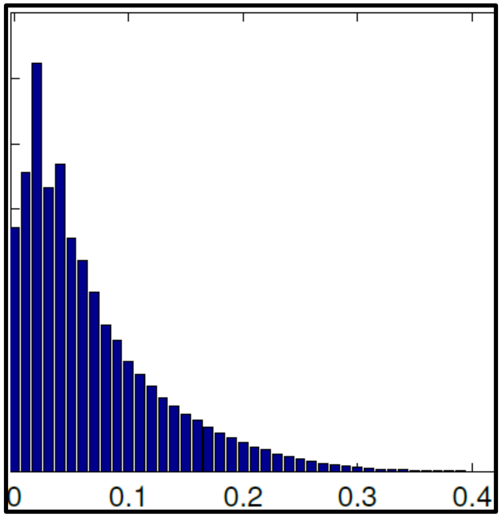
$$\frac{1}{2}N((N-1)^2 - N + 1)$$

- Have J experts place BB K-times -> N=JxK bounding boxes
- Collect ovelaps over 4 frames per sequence.
- All overlaps are examples of negligibly small difference
- Average can be taken as the threshold γ.

Estimation of practical difference thresholds



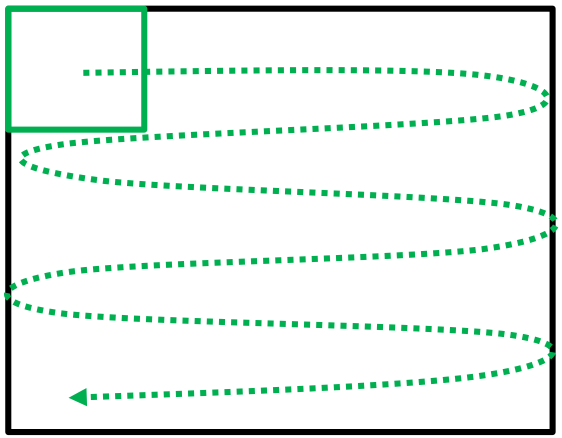
Distribution from all sequences



Examples of expert annotations

VOT2014 Speed measurement

- Reduce the hardware bias in reporting tracking speed.
- Approach: The VOT2014 speed benchmark

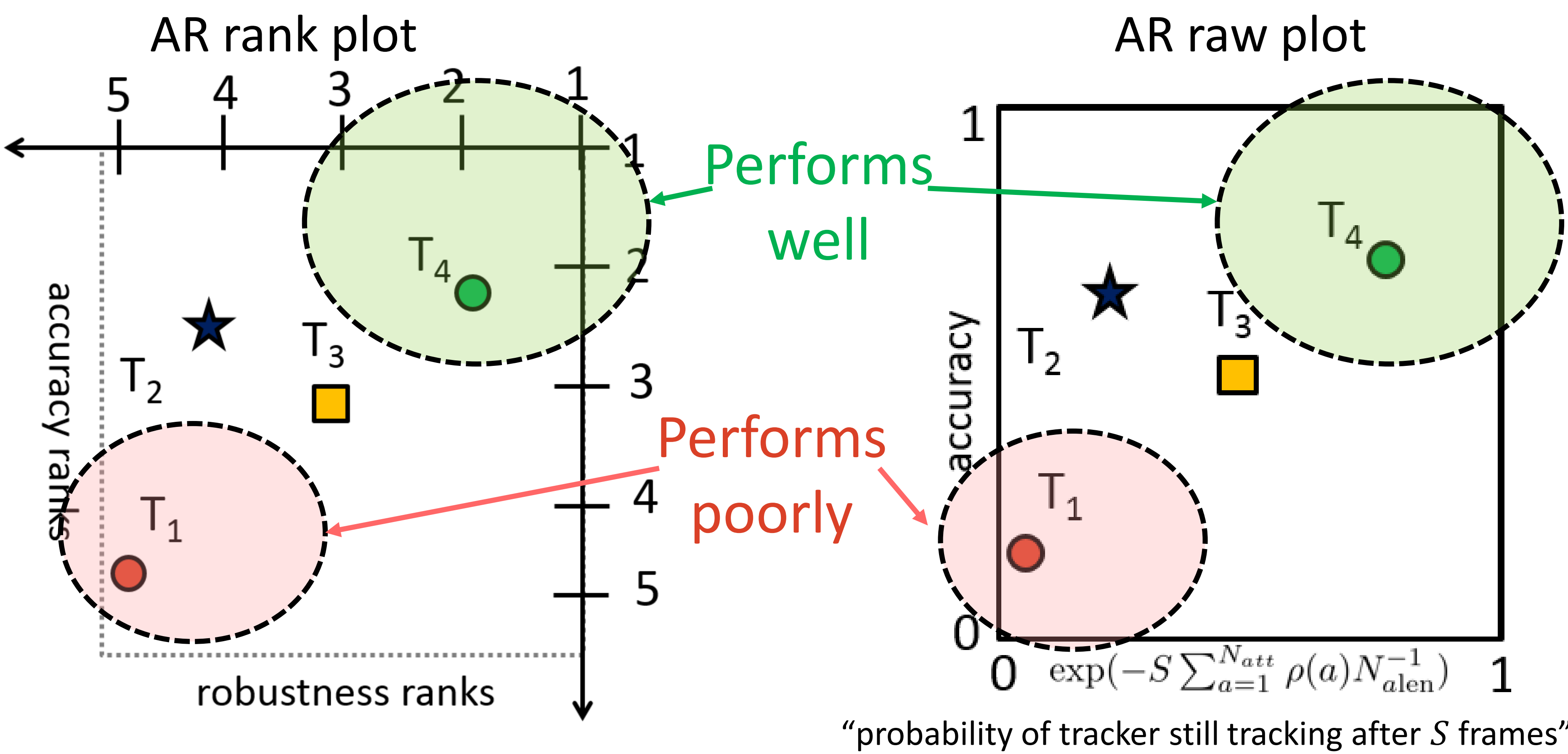


600x600 image
Max operation in 30x30 window
Apply this filter to all pixels
Measure the time for filtering

- Divide tracking time with time required to perform the filtering operation
- A new speed unit: Equivalent Filter Operations (EFO)

Visualizing the accuracy/robustness

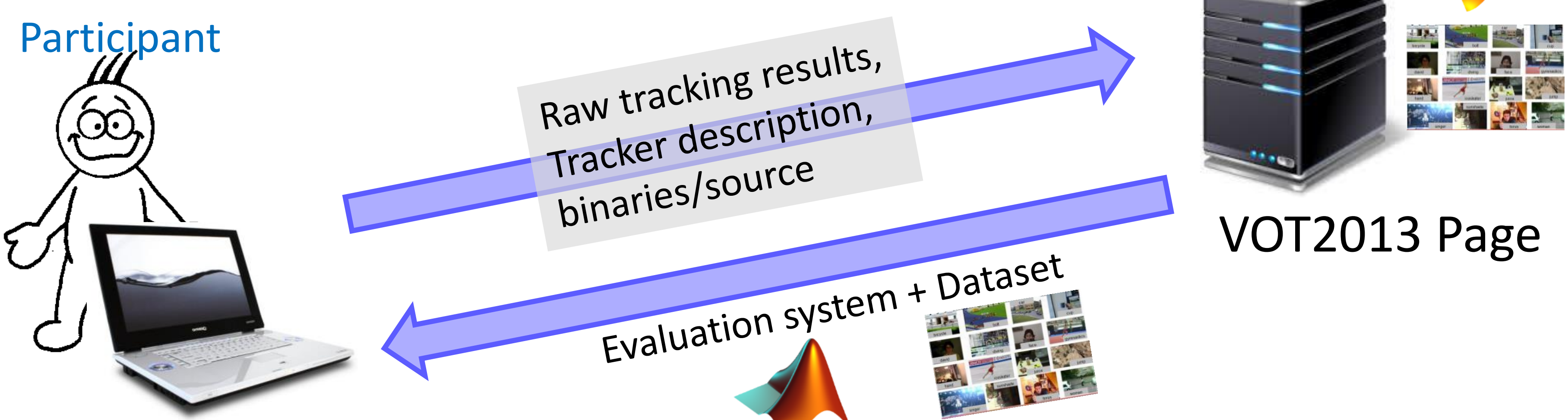
- AR rank plots as proposed in VOT2013
- AR raw plots as proposed by [Čehovin et al. 2014]



CHALLENGE PARTICIPATION AND SUBMITTED TRACKERS

VOT2014 Challenge: participation

- Participants would download the evaluation kit:
  - Evaluation system + Dataset
- Integrate their tracker into the evaluation system
- Predefined set of experiments automatically performed – submit the results back
- Required to submit binaries/source
- Required to outperform a NCC tracker



38 trackers tested!

33 entries from various authors + 5 baselines from VOT2014 committee = 38 trackers.

ABS	Possegger et al.	VOT 2014
ACAT	Qin et al.	CVPR 2014
ACT	Danelljan et al.	CVPR 2014
aStruck	Lukezic et al.	VOT 2014
BDF	Mareska et al.	VOT 2014
CMT	Nebehay et al.	VOT 2014
CT	Zhang et al.	ECCV2012
DGT	Wen et al.	ACCV 2012
DSST	Danelljan et al.	BMVC2014
DynMS	Oven et al	VOT 2014
eASMS	Vojir et al.	VOT 2014
EDFT	Felsberg	VOT 2013
MCT	Duffner et al.	VOT 2014
FoT	Vojir et al.	CVWW2011
FRT	Adam et al.	CVPR2006
SAMF	Li and Zhu	VOT 2014
SIR	Pangersic	VOT 2014
VTDMG	Moo Yi et al.	IVCNZ 2012

FSDT	Li et al.	VOT 2014
HMM-TxD	Vojir et al.	VOT 2014
IIVTv2	Moo Yi et al.	ICCV 2013
IPRT	Choi	VOT 2014
IMPNCC	Dimitriev	VOT 2014
IVT	Ross et al.	IJCV2008
KCF	Henriques et al.	TPAMI 2014
LGT	Cehovin et al.	VOT 2014
LT-FLO	Lebeda et al.	ACCV 2012
MatFlow	Mareska et al.	VOT 2014
Matrioska	Mareska et al.	ICIAP 2013
MIL	Babenko et al.	TPAMI2011
OGT	Nam et al.	VOT 2014
PLT13	Heng et al.	VOT 2013
PLT14	Heng et al.	VOT 2014
PT+	Duffner et al.	VOT 2014
qwsEDFT	Öfjäll et al.	VOT 2014
Struck	Hare et al.	ICCV 2011
TStruck	Hare et al.	VOT 2014

Tested trackers: rough categorization

Very diverse set of entries:

- Keypoint-based
  (CMT, IIVTv2, Matrioska, MatFlow)
- General part-based
  (LT-FLO, PT+, LGT, OGT, DGT, ABS)
- Global generative-model-based
  (EDFT, qwsEDFT, VTDGM, aSMS, IMPNCC,SIR-PF, IPRT,CT,IVT,HMM-TxD,DynMS)
- Discriminative models – single part
  (MCT, MIL, FSDT)
- Discriminative regression-based techniques
  (Struck, aStruck, ThunderStruck, PLT_{13},PLT_{14}, KCF, ACT, DSST)
- Combinations of multiple trackers
  (FoT, BDF, FRT, HMM-TxD, DynMS)

EXPERIMENTS AND RESULTS

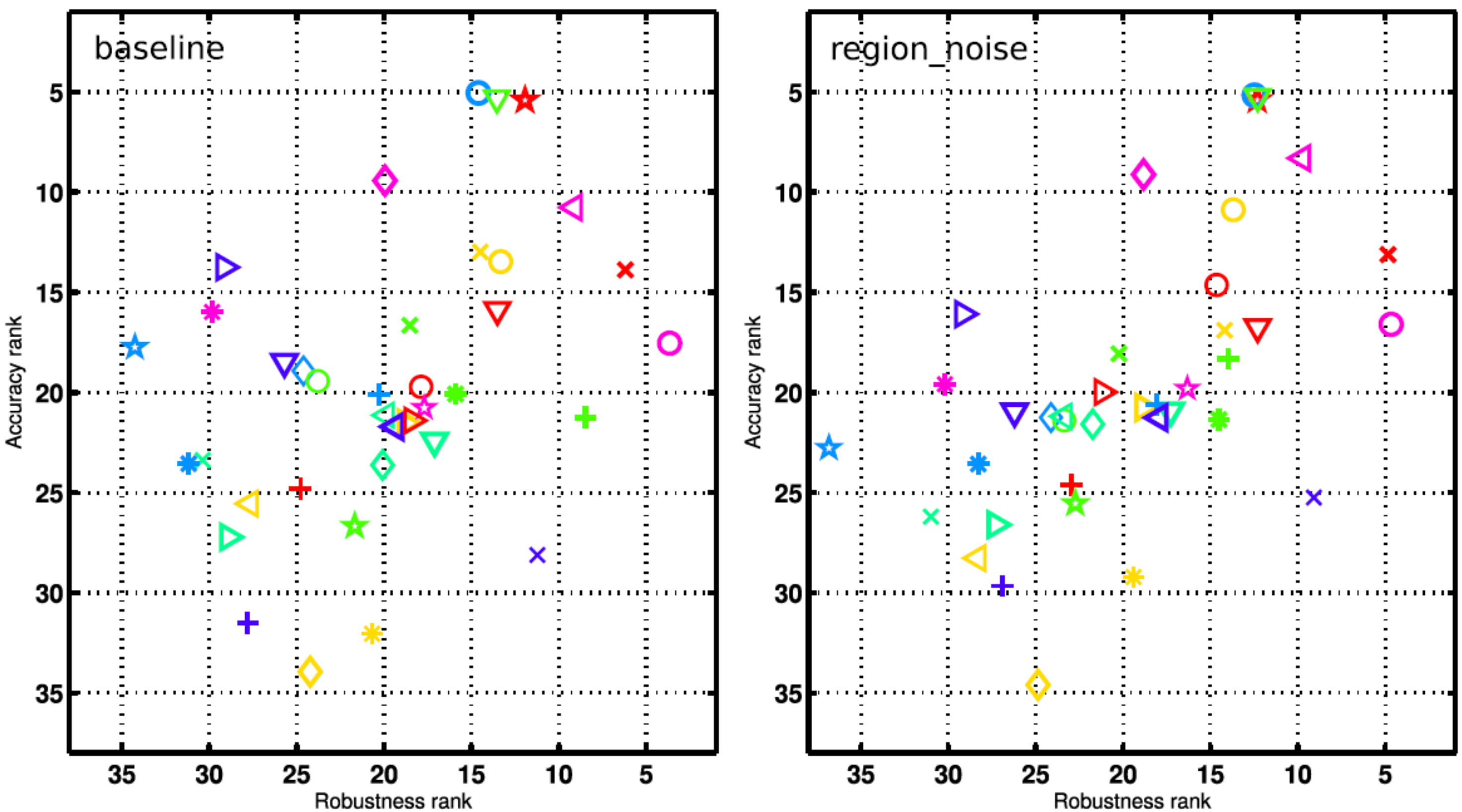
VOT2014 Experiments

- Experiment 1– Baseline:
  - 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.
- Each tracker run 15 times on each sequence to obtain a better statistic on its performance.
- Reinitialization threshold was 0.

Results: Experiments 1, 2

- Top-performing by averaging two experiments:

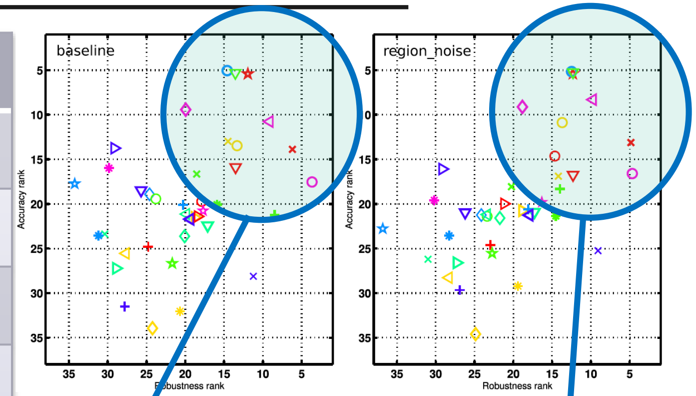
DSST, SAMF, KCF, DGT, PLT_{14}, PLT_{13}



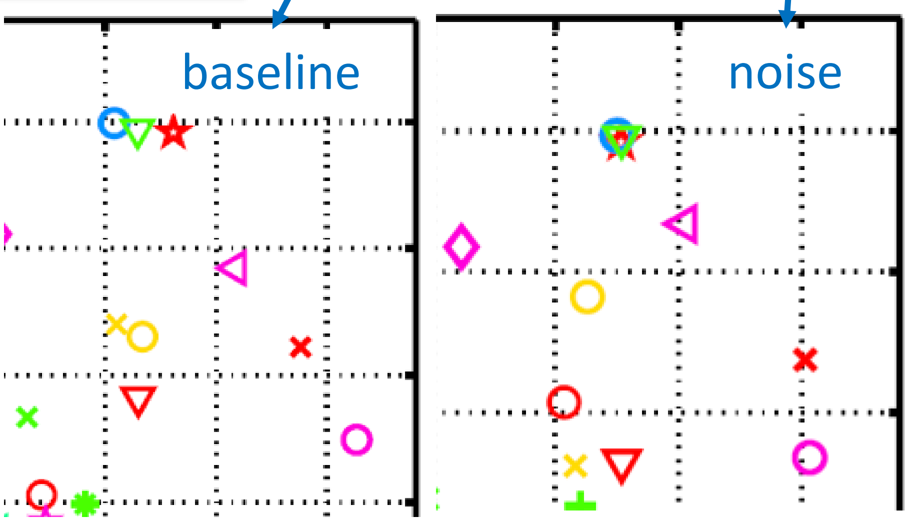
DSST*	8.77
SAMF*	9.10
KCF*	9.33
DGT	9.48
PLT_14*	9.51
PLT_13	10.62
eASMS*	12.85
HMM-TxD*	14.33
MCT	14.61
ACAT	14.65
MatFlow	15.51
ABS	16.72
ACT	17.97
qwsEDFT	18.37
LGT*	18.42
VTDMG	18.65
BDF	19.44
Struck	19.77
DynMS*	19.97
ThunderStruck	20.06
aStruck*	20.24
Matrioska	21.40
SIR-PF	21.76
EDFT	22.00
OGT	22.04
CMT*	22.23
FoT*	22.84
LT-FLO	23.90
IPRT	24.16
IIVTv2	24.29
PT+	25.34
FSDT	26.65
IMPNCC	27.45
IVT*	27.51
FRT*	27.74
NCC*	27.90
CT*	28.98
MIL*	29.41

Results: Experiments 1, 2

Tracker	Features	Scale	Visual model
DSST*	HoG+intensity	Yes	Discr. correl. Filtr
SAMF	HoG+colornames	Yes	Discr. correl. Filtr
KCF	HoG	Yes	Discr. correl. Filtr
DGT	Superpixels + color	Yes	Part-based
PLT_{14}	Color, intensity, derivs.	Yes	Discr. Regression
PLT_{13}	Color, intensity, derivs.	No	Discr. Regression



- Tight cluster (DSST,SAMF,KCF)
- Not-so tight (PLT13,PLT14)
- DGT somewhere in the middle

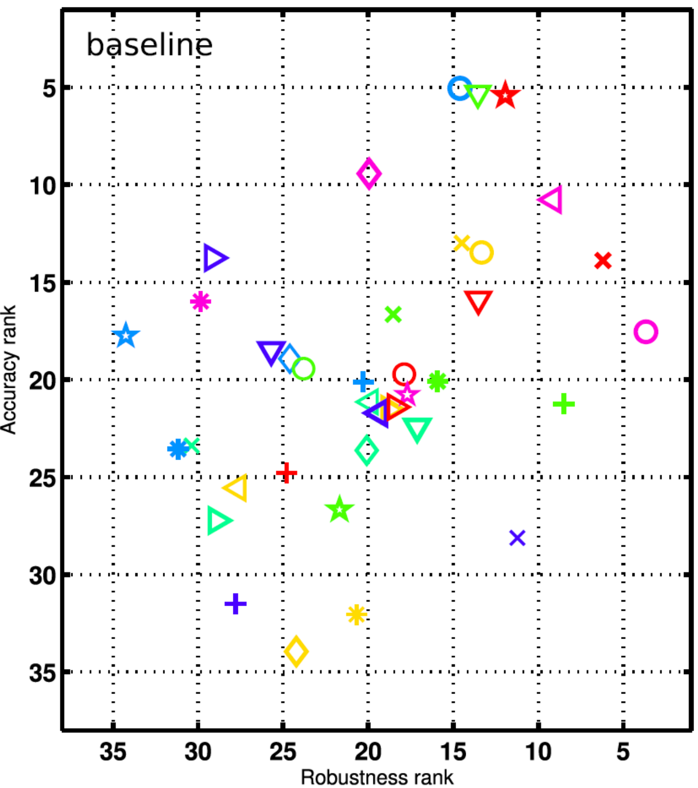


*Danelljan, M., Hager, G., Khan, F.S., Felsberg, M.: Accurate scale estimation for robust visual tracking. BMVC2014, (Talk today at 11:00)

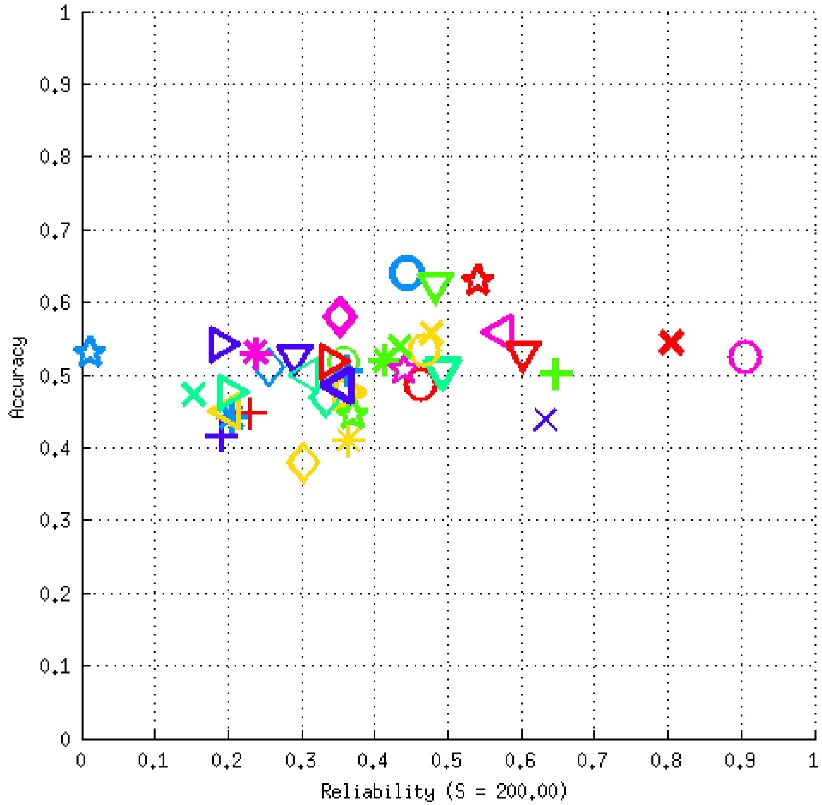
Results: Baseline experiment

- AR-rank plots vs raw AR plots

AR- rank plot



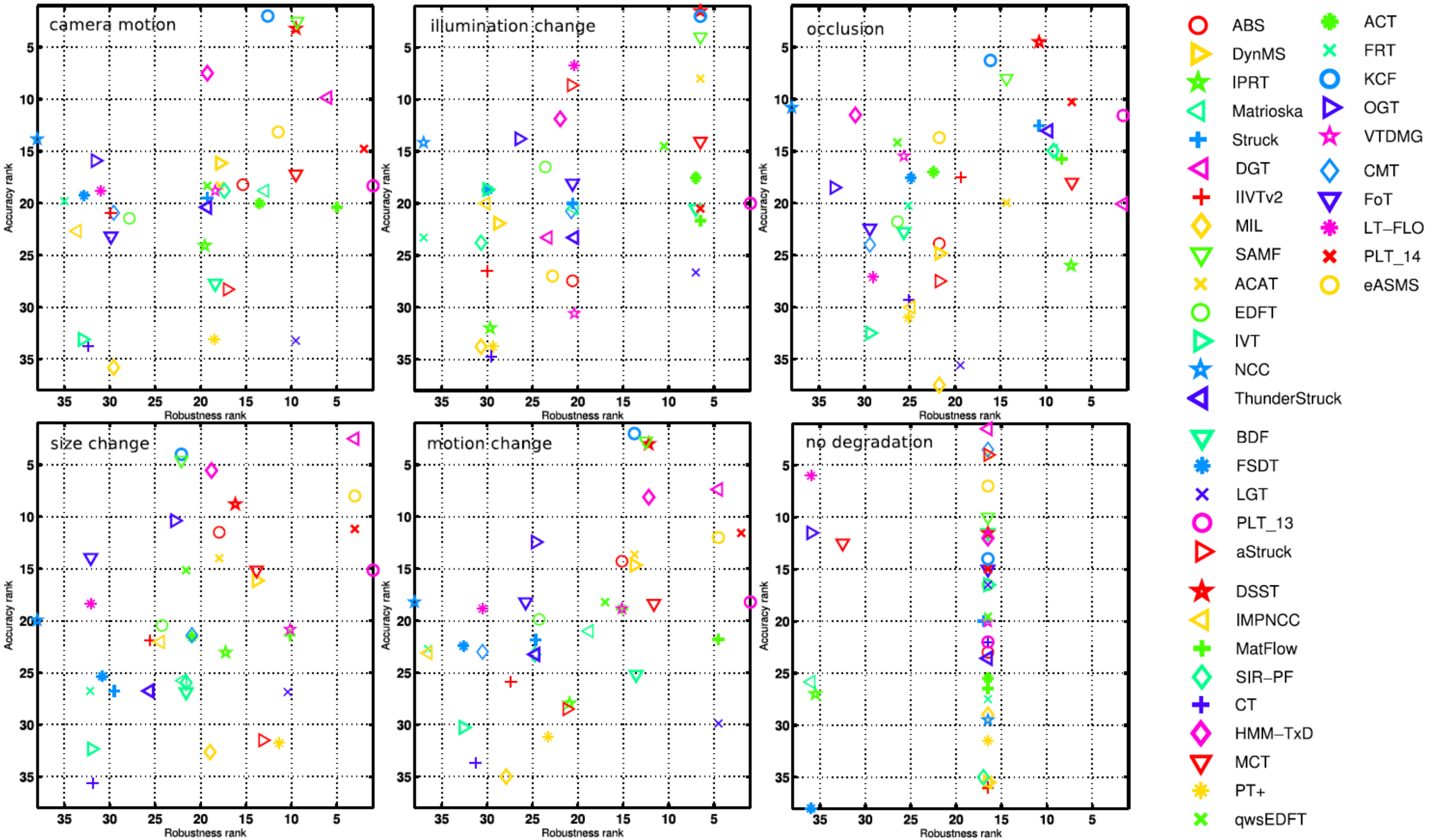
AR- raw plot
[Čehovin et al. 2014]



DSST*	8.77
SAMF*	9.10
KCF*	9.33
DGT	9.48
PLT_14*	9.51
PLT_13	10.62
eASMS*	12.85
HMM-TxD*	14.33
MCT	14.61
ACAT	14.65
MatFlow	15.51
ABS	16.72
ACT	17.97
qwsEDFT	18.37
LGT*	18.42
VTDMG	18.65
BDF	19.44
Struck	19.77
DynMS*	19.97
ThunderStruck	20.06
aStruck*	20.24
Matrioska	21.40
SIR-PF	21.76
EDFT	22.00
OGT	22.04
CMT*	22.23
FoT*	22.84
LT-FLO	23.90
IPRT	24.16
IIVTv2	24.29
PT+	25.34
FSDT	26.65
IMPNCC	27.45
IVT*	27.51
FRT*	27.74
NCC*	27.90
CT*	28.98
MIL*	29.41

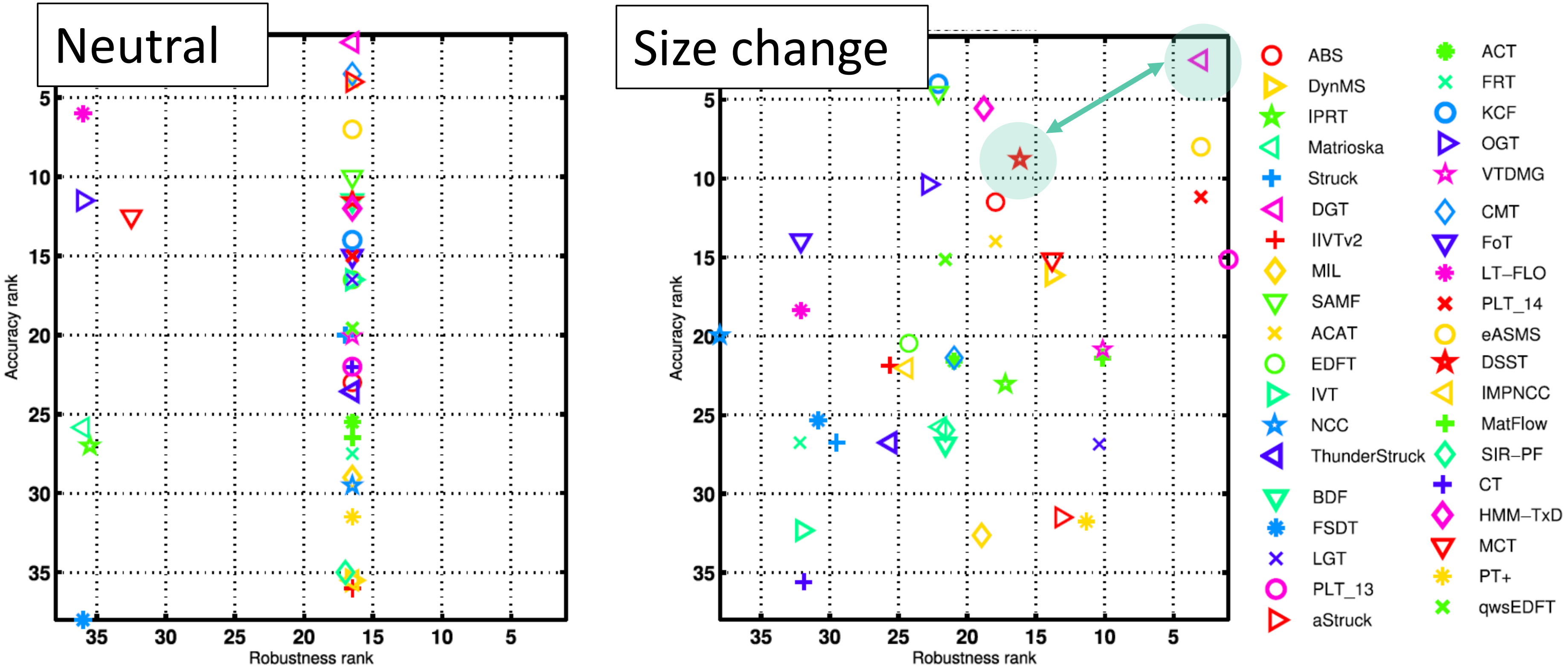
Performance w.r.t. attributes (Ex1)

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



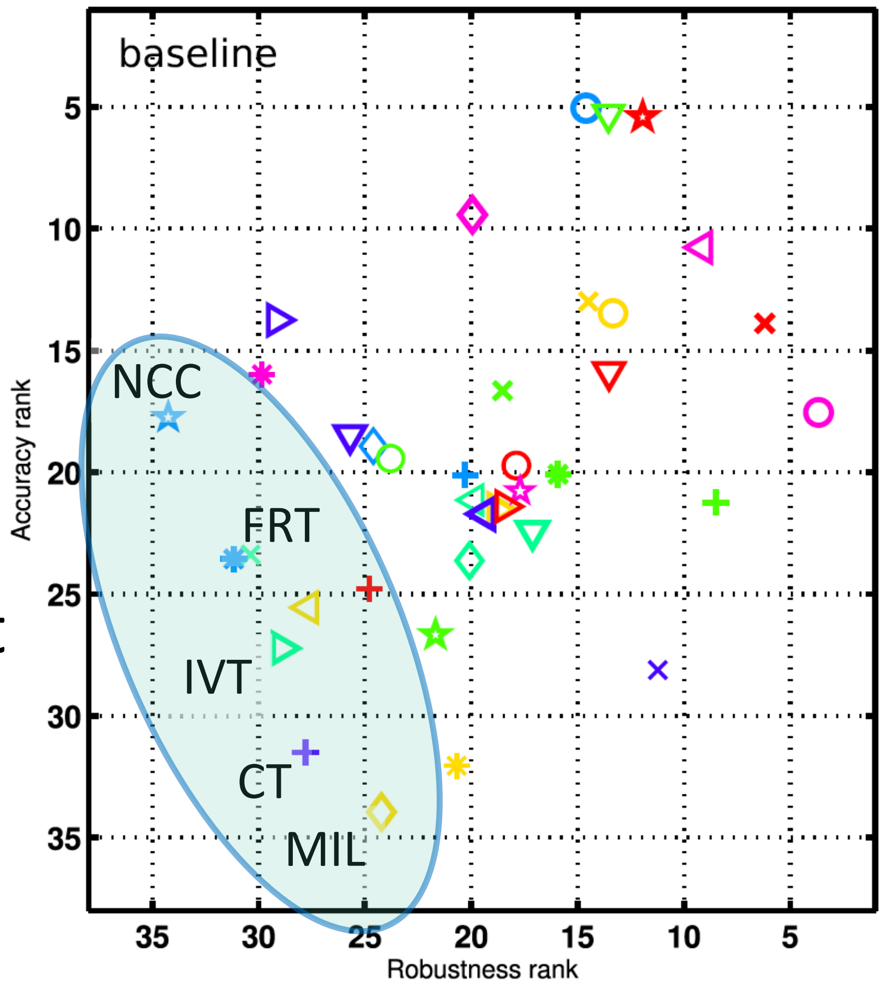
Performance w.r.t. attributes (Ex1)

- No degradation:
  - Most trackers equally robust, the difference only in accuracy (top DGT)
- Size change:
  - Significant switch in places (DGT and DSST)



VOT2014 trackers competitive

- Trackers that are often used as baseline SOTA: FRT,IVT,CT,MIL
- These occupy bottom-left part of AR rank plot.
- Same distribution apparent in noise experiment.



- Conclusion:
  Most tracker submitted to VOT2014 are competitive.

Tracking speed

- Fastest trackers:
  - FoT (C++) ~114 EFO
  - PLT_{13} (C++) ~75 EFO
- These were ere also the fastest in the VOT2013 challenge.

- For reference:
  - Type of tracker: NCC
  - Implementation: C++
  - Processor: Intel Core i5
  - Measured speed: 220 fps
  - EFO units: ~80 EFO

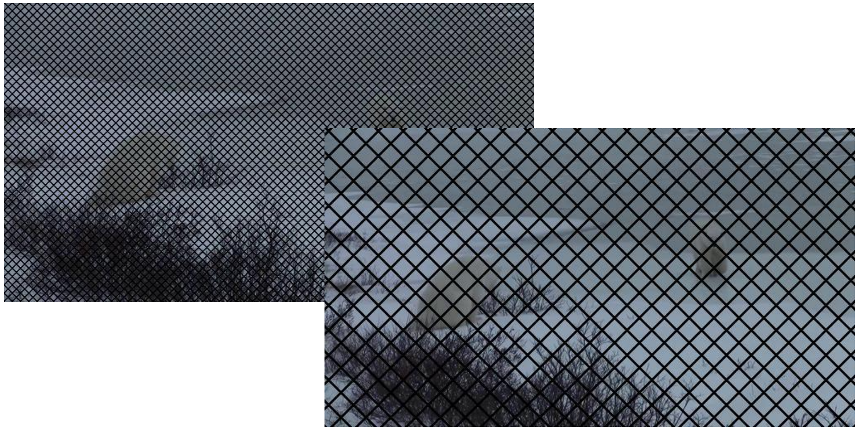
	Speed	Impl.
DSST*	7.66	Matlab & Mex
SAMF*	1.69	Matlab & Mex
KCF*	24.23	Matlab & Mex
DGT	0.23	C++
PLT_14*	62.68	C++
PLT_13	75.92	C++
eASMS*	13.08	C++
HMM-TxD*	2.08	C++
MCT	1.45	C, C++
ACAT	3.24	unknown
MatFlow	19.08	C++
ABS	0.62	Matlab & Mex
ACT	18.26	Matlab
qwsEDFT	3.88	Matlab
LGT*	1.23	Matlab & Mex
VTDMG	1.83	C++
BDF	46.82	C++
Struck	5.95	C++
DynMS*	3.21	Matlab & Mex
ThunderStruck	19.05	C++
aStruck*	3.58	C++
Matrioska	10.20	unknown
SIR-PF	2.55	Matlab & Mex
EDFT	4.18	Matlab
OGT	0.39	unknown
CMT*	2.51	Python, C++
FoT*	114.64	C++
LT-FLO	1.10	Matlab
IPRT	14.69	C, C++
IIVTv2	3.67	C++
PT+	49.89	C++
FSDT	1.47	C++
IMPNCC	8.37	Matlab
IVT*	2.35	Matlab & Mex
FRT*	3.09	C++
NCC*	6.88	Matlab
CT*	6.29	C++
MIL*	1.94	C++

Additional VOT2014 experiments

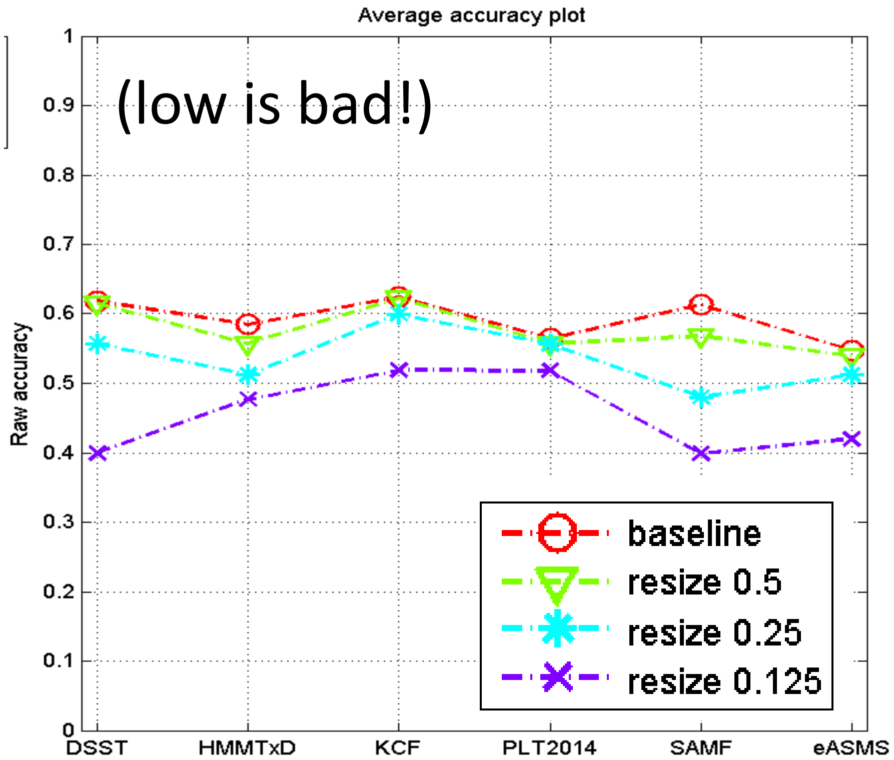
- Performed 2 variations of the Experiment 1 with six of the top-performing trackers
  (DSST, KCF, SAMF, PLT14, eASMS, HMMTxD)

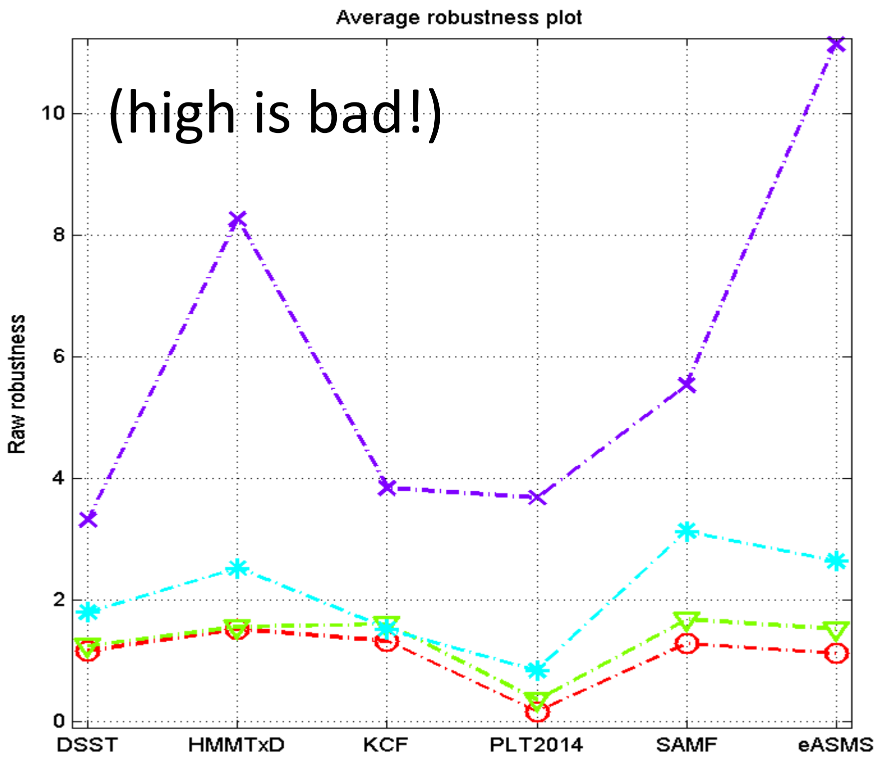
1. Sensitivity to object size:
   - Resize images by factor 0.5x, 0.25x, 0.125x.

2. Sensitivity to occlusion:
   - Place artificial static occluders in frames.



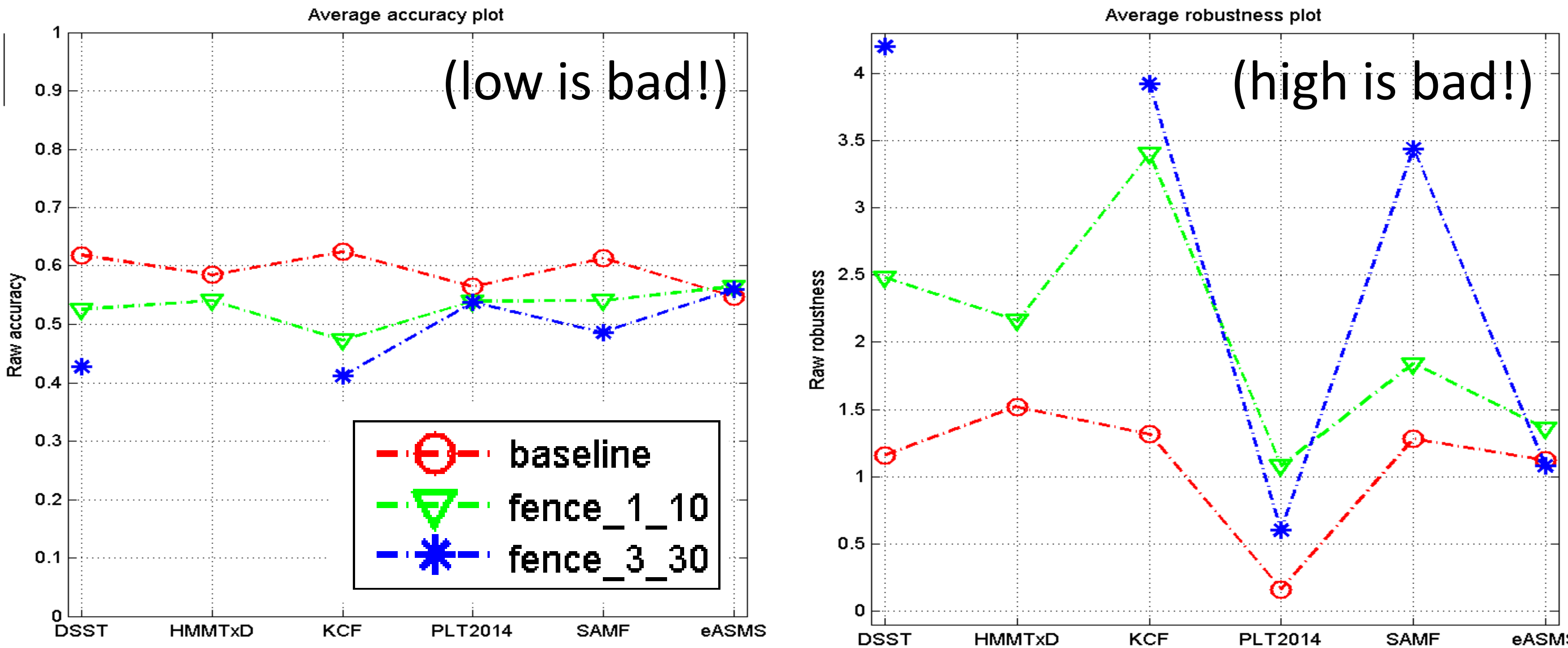
Resize experiment





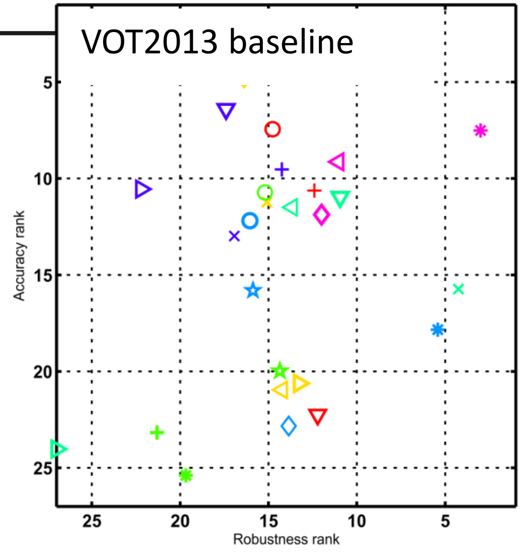
- PLT14 accuracy/robustness stable across scales
- Biggest drop for extreme resize (8x)

Occluder experiment



- PLT & eASMS least affected by the occluder
- Most significant drop in performance for correlation-based trackers (top three in VOT2014)

VOT2014 trackers on VOT2013



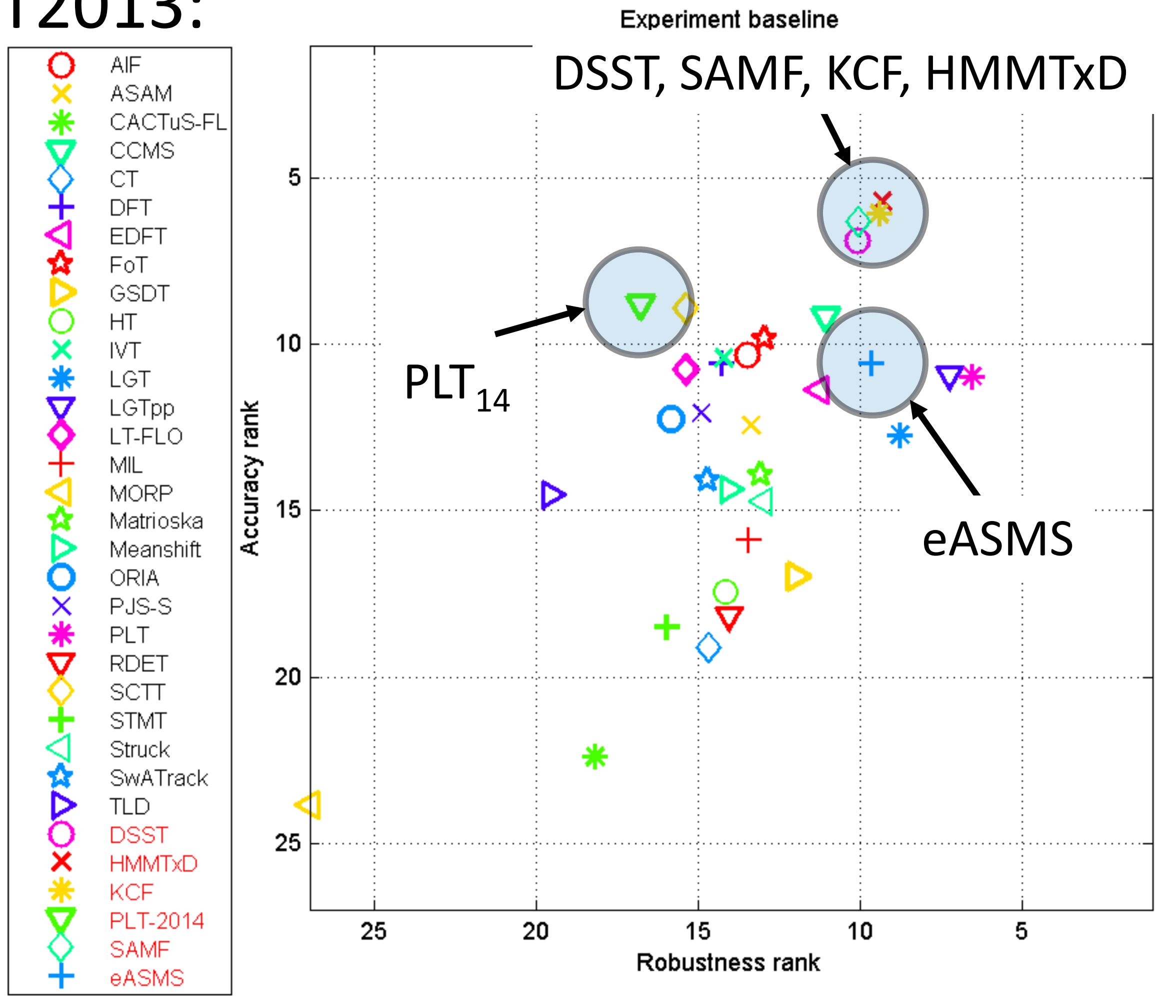
"Where in the VOT2013 AR plots are the top VOT2014 trackers positioned?"

- Approach:
  - Keep the rank positions of the VOT2013 trackers unchanged
  - Allow direct comparison to the VOT2013 AR rank plot
- If tracker performs better than best tracker, it gets rank 0.5
- If tracker performs better than T1, but poorer than T2, it gets the middle rank

VOT2014 trackers on VOT2013

- Trackers DSST, HMMTxD, KCF, PLT14, SAMF, eASMS positioned in VOT2013:

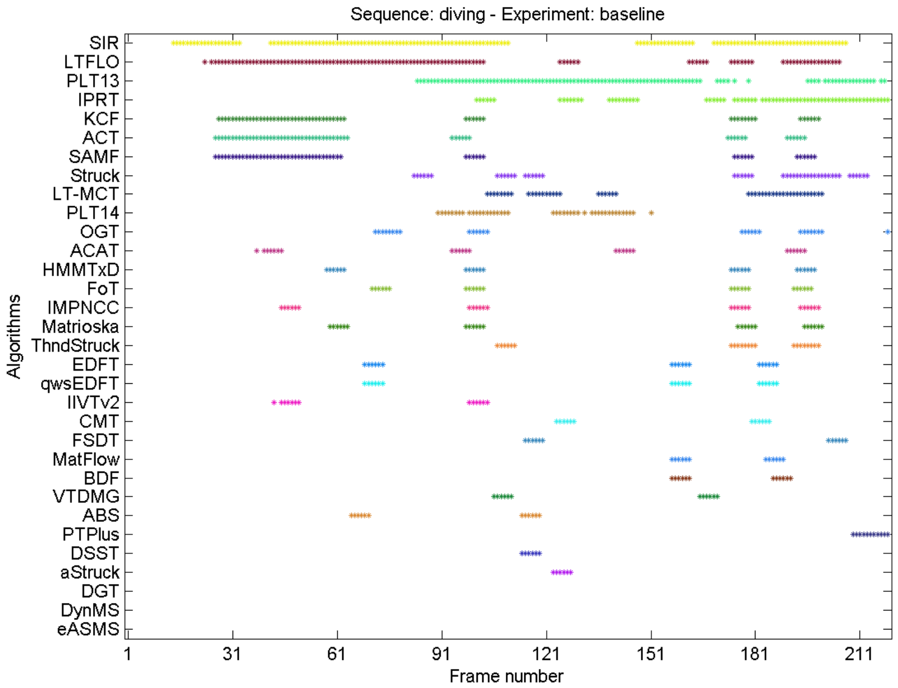
State-of-the-art advanced!



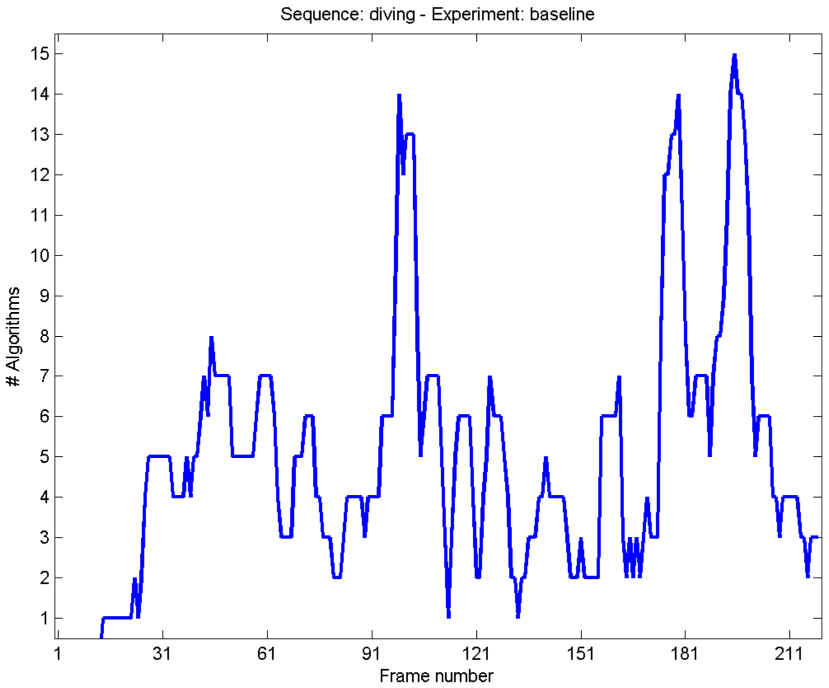
Sequence ranking

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

Failure frames for diving



How many trackers fail per frame?



Sequence ranking

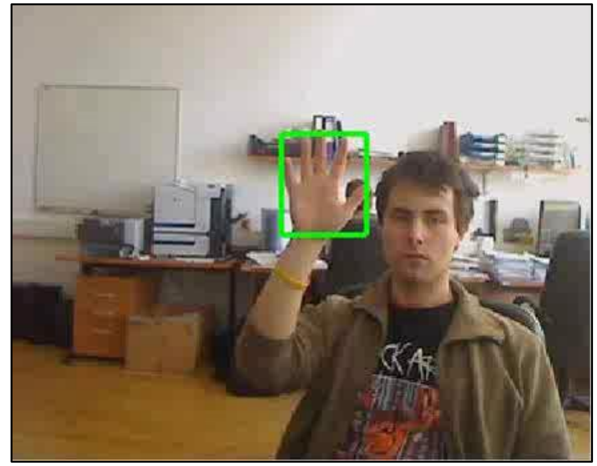
- Challenging: motocross, hand2, diving, fish2, bolt
- Intermediate: hand1, fish1, fernando, gymnastics, torus, Skating
- Easiest: Surfing, polarbear

Sequence	Baseline (Avg)
motocross	5,92
hand2	5,65
diving	4,85
fish2	4,59
bolt	4,14
hand1	3,23
fish1	2,94
fernando	2,78
gymnastics	2,59
torus	2,26
skating	2,12
trellis	1,58
basketball	1,43
tunnel	1,27
sunshade	1,24
jogging	1,12
woman	1,05
bicycle	0,75
david	0,60
ball	0,47
sphere	0,41
car	0,25
drunk	0,11
surfing	0,04
polarbear	0,00

Sequence ranking: Challenging

motocross
(camera and object motion + size change)

hand2
(object motion and size change)



diving
(camera motion at the end, size change)

bolt
(camera motion, object motion)

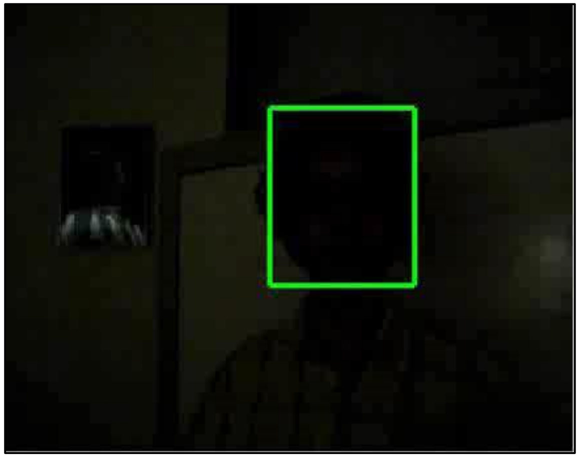
- motocross
- hand2
- diving
- fish2
- bolt
- hand1
- fish1
- fernando
- gymnastics
- torus
- skating
- trellis
- basketball
- tunnel
- sunshade
- jogging
- woman
- bicycle
- david
- ball
- sphere
- car
- drunk
- surfing
- polarbear

Sequence ranking: Less challenging

ball
(camera and object motion)

drunk
(artifacts)

david
(camera motion, illumination)



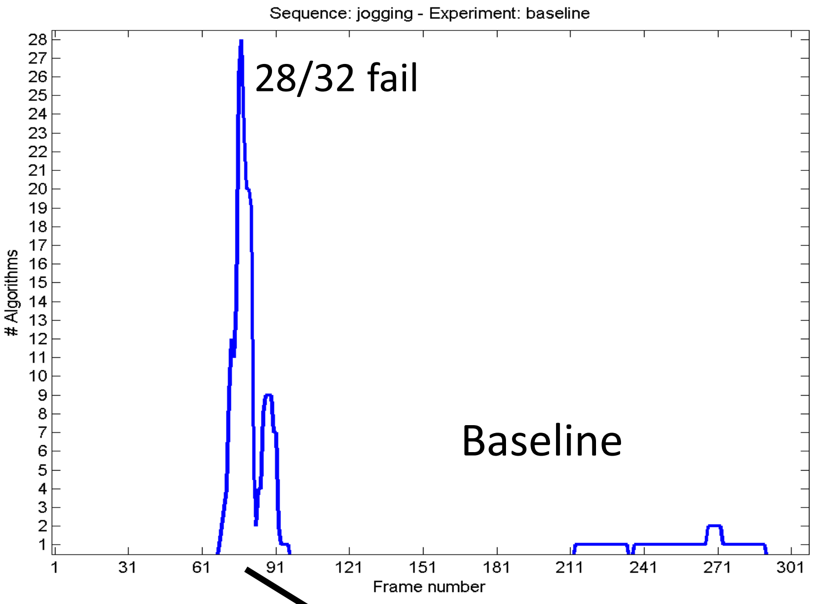
surfing
(camera motion, object motion)

motocross
hand2
diving
fish2
bolt
hand1
fish1
fernando
gymnastics
torus
skating
trellis
basketball
tunnel
sunshade
jogging
woman
bicycle
david
ball
sphere
car
drunk
surfing
polarbear

Sequence ranking: Locality

- Jogging: on average not challenging, but very challenging at particular frame span where almost all trackers fail

Less challenging: Jogging



Assumed cause: Occlusion!

- Locality: a sequence may be challenging only locally

VOT Summary: Results

- None of the trackers consistently outperformed all others by all measures
- The top-performing trackers included single-patch-based as well as part-based trackers.
- Robustness best for discriminative trackers, e.g., PLT_{13}
- Best tradeoff in accuracy and robustness achieved by correlation-based trackers
- Top VOT2014 trackers also top performing on VOT2013!

The VOT2014 online resources

Available at: http://www.votchallenge.net/vot2014

- This presentation and all papers
- Source code/binaries of some trackers
- Dataset + Evaluation kit
- Guidelines on how to evaluate your trackers on VOT2014 and produce graphs for your papers (directly comparable to 38 trackers!)

A new online portal for interactive analysis of results!

- Will be presented by its author, Luka Čehovin.

VOT2014 summary

- Results published in a 27 pages joint paper

The Visual Object Tracking VOT2014 challenge results

Matej Kristan[1], Roman Pflugfelder[2], Aleš Leonardis[3], Jiri Matas[4], Luka Čehovin[1], Georg Nebehay[2], Tomáš Vojíř[4], Gustavo Fernández[2], Alan Lukežič[1], Aleksandar Dimitriev[1], Alfredo Petrosino[5], Amir Saffari[6], Bo Li[7], Bohyung Han[8], CherKeng Heng[7], Christophe Garcia[9], Dominik Pangeršič[1], Gustav Häger[10], Fahad Shahbaz Khan[10], Franci Oven[1], Horst Possegger[11], Horst Bischof[11], Hyeonseob Nam[8], Jianke Zhu[12], JiJia Li[13], Jin Young Choi[14], Jin-Woo Choi[15], João F. Henriques[16], Joost van de Weijer[17], Jorge Batista[16], Karel Lebeda[18], Kristoffer Öfjäll[10], Kwang Moo Yi[19], Lei Qin[20], Longyin Wen[21], Mario Edoardo Maresca[5], Martin Danelljan[10], Michael Felsberg[10], Ming-Ming Cheng[22], Philip Torr[22], Qingming Huang[23], Richard Bowden[18],

Obvious Engineering Limited, United Kingdom

Winners of the VOT2014 challenge:

DSST by Martin Danelljan, Gustav Hager, Fahad Khan, and Michael Felsberg

Reference: *Danelljan, M., Hager, G., Khan, F.S., Felsberg, M. Accurate scale estimation for robust visual tracking. BMVC2014,*

Presentation: on VOT2014 today at 11:00

Abstract. The Visual Object Tracking challenge 2014, VOT2014, aims at comparing short-term single-object visual trackers that do not apply pre-learned models of object appearance. Results of 37 trackers are

Thanks

- The VOT2014 committee

- Everyone who participated!

et al.: Alan Lukežič (Ljubljana University), Aleksandar Dimitriev (Ljubljana University), Alfredo Petrosino (Parthenope University of Naples), Amir Saffari (Affectv Limited), Bo Li (Panasonic R&D Center), Bohyung Han (POSTECH), CherKeng Heng (Panasonic R&D Center), Christophe Garcia (LIRIS), Dominik Pangeršič (Ljubljana University), Gustav Häger (Linköping University), Fahad Shahbaz Khan (Linköping University), Franci Oven (Ljubljana University), Horst Possegger (Graz University of Technology), Horst Bischof (Graz University of Technology), Hyeonseob Nam (POSTECH), Jianke Zhu (Zhejiang University), JiJia Li (Shanghai Jiao Tong University), Jin Young Choi (Seoul National University, ASRI), Jin-Woo Choi (Electronics and Telecommunications Research Institute, Daejeon), João F. Henriques (University of Coimbra), Joost van de Weijer (Universitat Autonoma de Barcelona), Jorge Batista (University of Coimbra), Karel Lebeda (University of Surrey), Kristoffer Öfjäll (Linköping University), Kwang Moo Yi (EPFL CVLab), Lei Qin (ICT CAS), Longyin Wen (Chinese Academy of Sciences), Mario Edoardo Maresca (Parthenope University of Naples), Martin Danelljan (Linköping University), Michael Felsberg (Linköping University), Ming-Ming Cheng (University of Oxford), Philip Torr (University of Oxford), Qingming Huang (Harbin Institute of Technology), Richard Bowden (University of Surrey), Sam Hare (Obvious Engineering Limited), Samantha YueYing Lim (Panasonic R&D Center), Seunghoon Hong (POSTECH), Shengcai Liao (Chinese Academy of Sciences), Simon Hadfield (University of Surrey), Stan Z. Li (Chinese Academy of Sciences), Stefan Duffner (LIRIS), Stuart Golodetz (University of Oxford), Thomas Mauthner (Graz University of Technology), Vibhav Vineet (University of Oxford), Weiyao Lin (Shanghai Jiao Tong University), Yang Li (Zhejiang University), Yuankai Qi (Harbin Institute of Technology), Zhen Lei (Chinese Academy of Sciences), ZhiHeng Niu (Panasonic R&D Center).

- Sponsor of VOT2014:

University *of Ljubljana*
Faculty *of Computer and*
Information Science

This work was supported in part by the following research programs and projects: Slovenian research agency projects J24284, J23607 and J2-2221, EU 7FP grant no 257906, CTU Project SGS13/142/OHK3/2T/13 and by the Technology Agency of the Czech Republic project TE01020415 (V3C{ Visual Computing Competence Center).

Note

- Some slides were modified after the VOT2014 presentation to reflect further details in the evaluation results.