Benchmarking the State of the Art in Visual Tracking

吴毅
https://sites.google.com/site/wuyi2018/
江苏省大数据分析技术重点实验室

2014.10.15

Tracking in Computer Vision

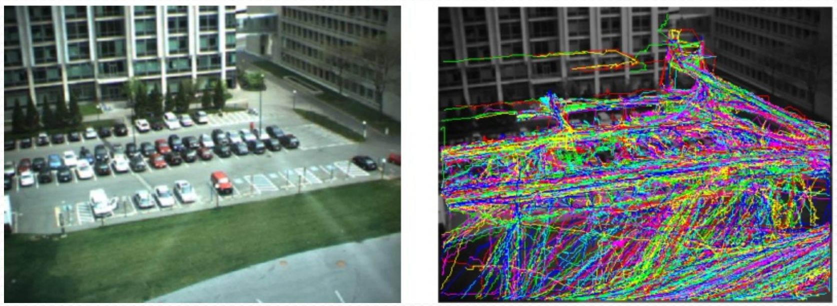
Initialization in the 1st frame

……..

Estimated states in the N-th frame

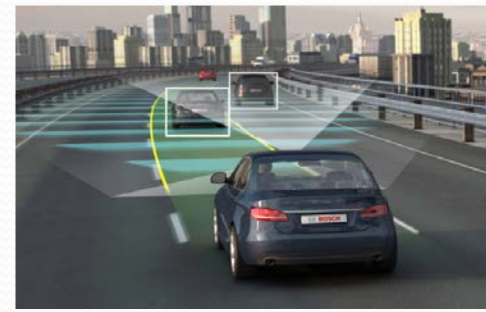
- A fundamental problem in computer vision
- A challenging and difficult task
- Numerous applications

Applications



Motion analysis

Surveillance



Autonomous robots/cars

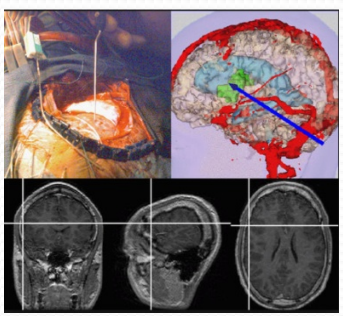
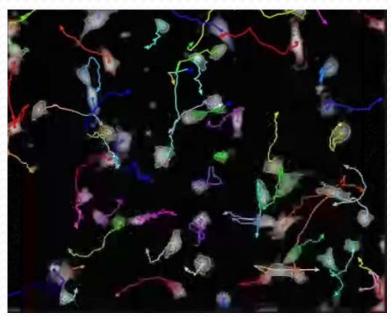


Image Guided Surgery



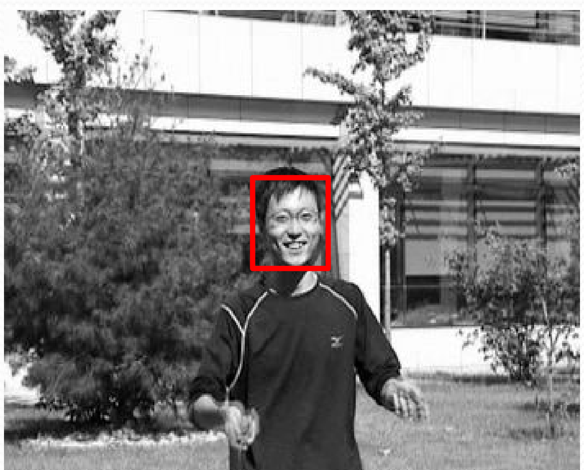
Biomedical image analysis

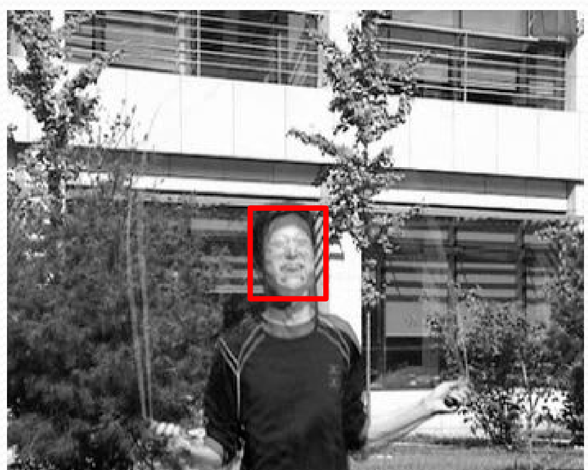
Human computer interaction

Tracking Challenges

deformation

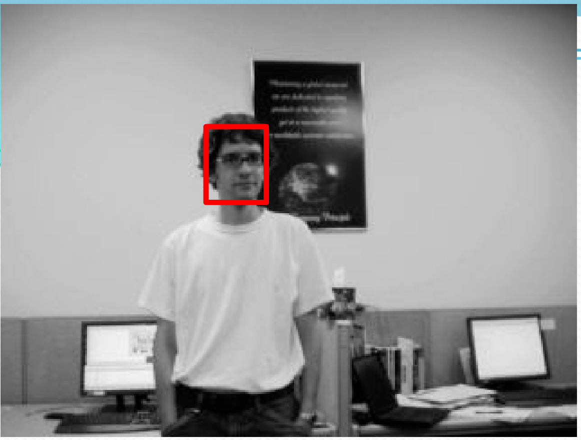
illumination variation

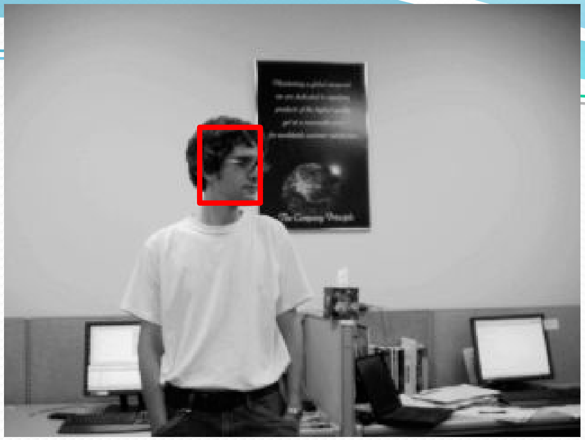




blur & fast motion

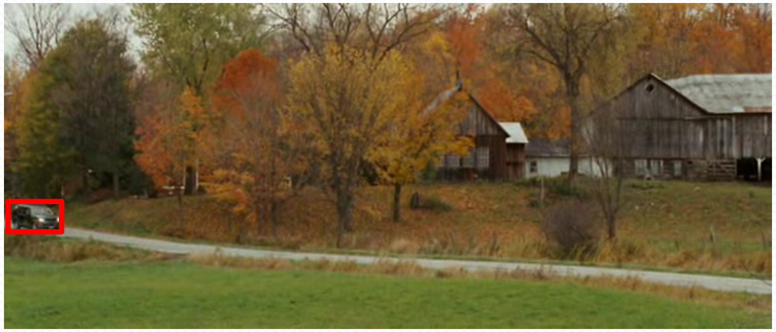
background clutter



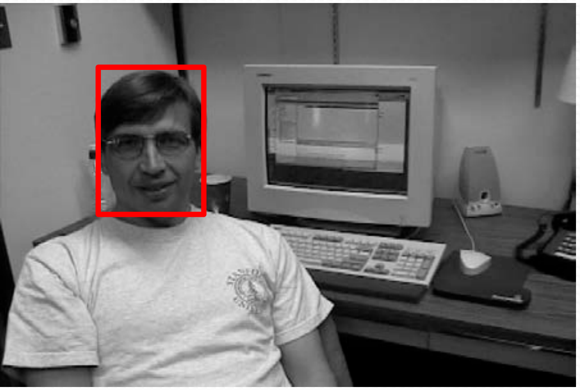


out-of-plane rotation

in-plane rotation



scale variation



occlusion

out-of-view

Tracking techniques

- Object representation
- Searching mechanism
- Model update
- Advanced tracking techniques
  - Mixture model
    - Observation
    - Dynamic model
    - Tracker
  - Context information
  - Re-detection

Object representation

- Holistic feature
  - Intensity template
  - Color/intensity histograms
    - **MS**: D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," T-PAMI, vol. 25, no. 5, pp. 564–577, 2003.
- Local feature
  - Histograms of Oriented Gradient (HOG)
  - Local Binary Patterns (LBP)
  - Fragments-based representation
    - **Frag**: A. Adam, E. Rivlin, and I. Shimshoni, "Robust Fragments-based Tracking using the Integral Histogram," in CVPR, 2006.

Object representation

- Generative model
  - Subspace representation
    - PCA
      - **IVT**: D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental Learning for Robust Visual Tracking," IJCV, 2008.
    - Sparse representation
      - X. Mei and H. Ling, "Robust Visual Tracking using L1 Minimization ," in *ICCV*, 2009.

Object representation

- Discriminative model
  - A binary classifier is on-line learned to discriminate the target from the background
  - Encoding the background information
  - Machine learning techniques
    - **SVM**: S. Avidan, “Support Vector Tracking,” T-PAMI, vol. 26, no. 8, pp. 1064–1072, 2004.
    - **Structured SVM**: S. Hare, A. Saffari, and P. H. S. Torr, “Struck: Structured output tracking with kernels,” in *ICCV*, 2011.
    - **Boosting**: S. Avidan, “Ensemble Tracking,” T-PAMI, vol. 29, no. 2, pp. 261–271, 2008.
    - **Online Boosting**: H. Grabner, M. Grabner, and H. Bischof, “Real-Time Tracking via On-line Boosting,” in BMVC, 2006.
    - **Online Multiple Instance Boosting**: B. Babenko, M.-H. Yang, and S. Belongie, “Visual Tracking with Online Multiple Instance Learning,” in CVPR, 2009.

Searching mechanism

- Deterministic method
  - Local optimum search
    - Lucas-Kanade
    - Mean Shift
  - Dense sampling
    - Learning based approach
- Stochastic method (Particle filter)
  - A flexible tracking framework
    - How to design the likelihood

Model update

- Adapting the model to the appearance variations of target
  - Template update
    - I. Matthews, T. Ishikawa, and S. Baker, “The Template Update Problem,” T-PAMI, 2004.
  - Incremental subspace learning
    - **IVT**: D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, “Incremental Learning for Robust Visual Tracking,” IJCV, 2008.
  - Online-learning of discriminative model
    - Online Boosting
    - Online SVM

Advanced tracking techniques

- Mixture of observations
  - Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, “Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Life Spans,” *T-PAMI*, vol. 30, no. 10, pp. 1728–1740, 2008.
  - B. Stenger, T. Woodley, and R. Cipolla, “Learning to Track with Multiple Observers,” in *CVPR*, 2009.
  - J. Kwon and K. M. Lee, “Visual Tracking Decomposition,” in *CVPR*, 2010.
- Mixture of dynamic models
  - J. Kwon and K. M. Lee, “Visual Tracking Decomposition,” in CVPR, 2010.

Advanced tracking techniques

- Mixture of trackers
  - B. Zhong, H. Yao, S. Chen, R. Ji, X. Yuan, S. Liu, and W. Gao, “Visual tracking via weakly supervised learning from multiple imperfect oracles,” in *CVPR*, 2010.
  - J. Santner, C. Leistner, A. Saffari, T. Pock, and H. Bischof, “PROST: Parallel Robust Online Simple Tracking,” in *CVPR*, 2010.
  - J. Kwon and K. M. Lee, “Tracking by Sampling Trackers,” in *ICCV*, 2011.
  - J. H. Yoon, D. Y. Kim, and K. Yoon, “Visual Tracking via Adaptive Tracker Selection with Multiple Features,” in *ECCV*, 2012.

Advanced tracking techniques

- Context information
  - T. B. Dinh, N. Vo, and G. Medioni, "Context tracker: Exploring supporters and distracters in unconstrained environments," in CVPR, 2011.
  - M. Yang, Y. Wu, and G. Hua, "Context-aware visual tracking," *T-PAMI*, pp. 1195–1209, 2008.
  - H. Grabner, J. Matas, L. V. Gool, and P. Cattin, "Tracking the Invisible: Learning Where the Object Might be," in *CVPR*, 2010.

Advanced tracking techniques

- Re-detection
  - H. Grabner, M. Grabner, and H. Bischof, "Real-Time Tracking via On-line Boosting," in BMVC, 2006.
  - H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised On-Line Boosting for Robust Tracking," in ECCV, 2008.
  - S. Stalder, H. Grabner, and L. van Gool, "Beyond Semi-Supervised Tracking: Tracking Should Be as Simple as Detection, but not Simpler than Recognition," in ICCV Workshop, 2009.
  - Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints," CVPR, 2010.
  - T. B. Dinh, N. Vo, and G. Medioni, "Context tracker: Exploring supporters and distracters in unconstrained environments," in CVPR, 2011.

Problems in Tracking

- Weakness of most proposed trackers is not clear
- Bias in the performance comparison
  - Different ground truth bounding boxes
  - Different initializations

Problems in Tracking

- Weakness of most proposed trackers is not clear
- Bias in the performance comparison
  - Different ground truth bounding boxes
  - Different initializations

Which tracker is more robust?

Problems in Tracking

- Weakness of most proposed trackers is not clear
- Bias in the performance comparison
  - Different ground truth bounding boxes
  - Different initializations

A Benchmark is urgent for tracking!

Tracking Benchmark

- Evaluated Trackers and Dataset
- Evaluation Methodology
- Evaluation Results
- Concluding Remarks

Yi Wu
Nanjing University of Information Science and Technology, China

Jongwoo Lim
Hanyang university, Korea

MingHsuan Yang
University of California, Merced, USA

[1] Y. Wu, J. Lim, and M.-H. Yang, "Online Object Tracking: A Benchmark," in *CVPR*, 2013.
[2] Y. Wu, J. Lim, and M.-H. Yang, "Object Tracking Benchmark," T-PAMI, 2014, in press.

Evaluated Trackers

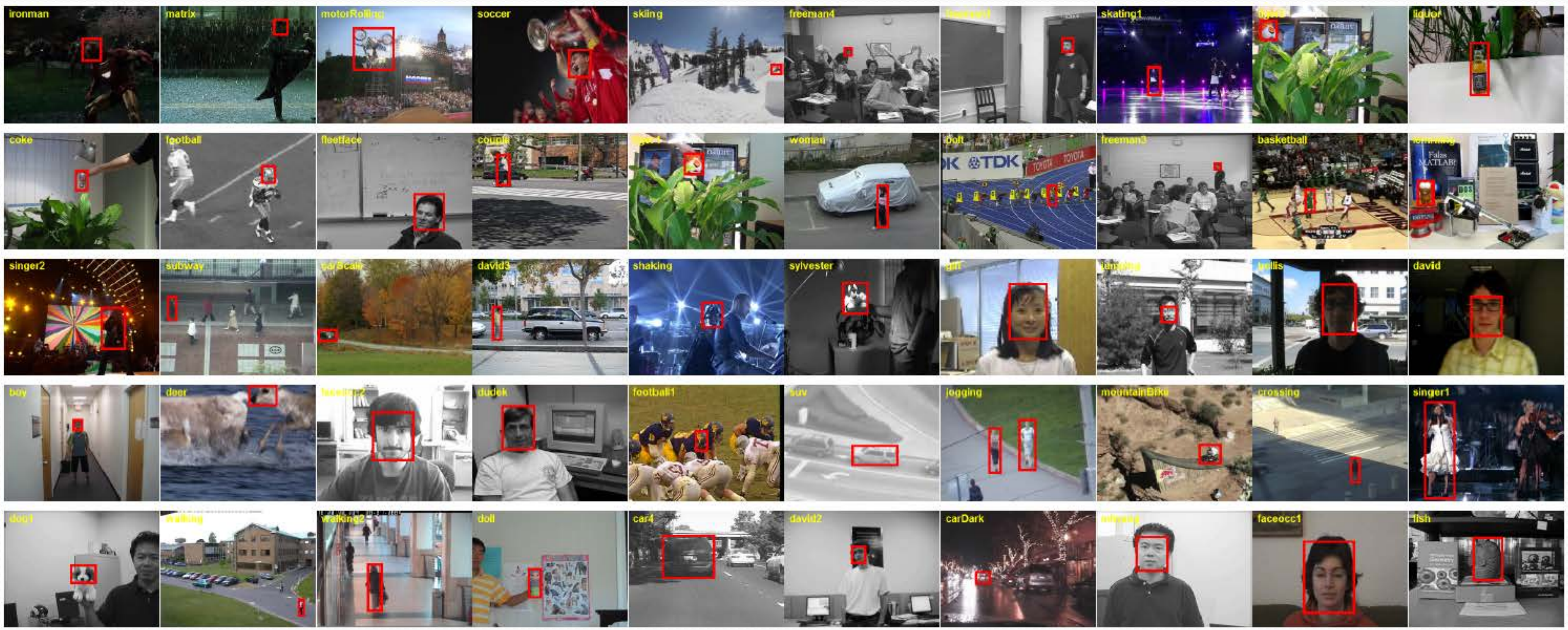
NAME	CODE	REFERENCE
CPF	CPF	P. Pe´rez, C. Hue, J. Vermaak, and M. Gangnet. Color-Based Probabilistic Tracking. In ECCV, 2002.
KMS	KMS	D. Comaniciu, V. Ramesh, and P. Meer. Kernel-Based Object Tracking. PAMI, 25(5):564–577, 2003.
SMS	SMS	R. Collins. Mean-shift Blob Tracking through Scale Space. In CVPR, 2003.
VR-V	VIVID/VR	R. T. Collins, Y. Liu, and M. Leordeanu. Online Selection of Discriminative Tracking Features. PAMI, 27(10):1631–1643, 2005. [www] * We also evaluated four other trackers included in the VIVID tracker suite. (**PD-V**, **RS-V**, **MS-V**, and **TM-V**).
Frag	Frag	A. Adam, E. Rivlin, and I. Shimshoni. Robust Fragments-based Tracking using the Integral Histogram. In CVPR, 2006. [www]
OAB	OAB	H. Grabner, M. Grabner, and H. Bischof. Real-Time Tracking via On-line Boosting. In BMVC, 2006. [www]
IVT	IVT	D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental Learning for Robust Visual Tracking. IJCV, 77(1):125–141, 2008. [www]
SemiT	SBT	H. Grabner, C. Leistner, and H. Bischof. Semi-supervised On-Line Boosting for Robust Tracking. In ECCV, 2008. [www]
MIL	MIL	B. Babenko, M.-H. Yang, and S. Belongie. Visual Tracking with Online Multiple Instance Learning. In CVPR, 2009. [www]
BSBT	BSBT	S. Stalder, H. Grabner, and L. van Gool. Beyond Semi-Supervised Tracking: Tracking Should Be as Simple as Detection, but not Simpler than Recognition. In ICCV Workshop, 2009. [www]
TLD	TLD	Z. Kalal, J. Matas, and K. Mikolajczyk. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. In CVPR, 2010. [www]
VTD	–	J. Kwon and K. M. Lee. Visual Tracking Decomposition. In CVPR, 2010. [www]
CXT	CXT	T. B. Dinh, N. Vo, and G. Medioni. Context Tracker: Exploring supporters and distracters in unconstrained environments. In CVPR, 2011. [www]
LSK	LSK	B. Liu, J. Huang, L. Yang, and C. Kulikowsk. Robust Tracking using Local Sparse Appearance Model and K-Selection. In CVPR, 2011. [www]
Struck	Struck	S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured Output Tracking with Kernels. In ICCV, 2011. [www]
VTS	–	J. Kwon and K. M. Lee. Tracking by Sampling Trackers. In ICCV, 2011. [www]
ASLA	ASLA	X. Jia, H. Lu, and M.-H. Yang. Visual Tracking via Adaptive Structural Local Sparse Appearance Model. In CVPR, 2012. [www]
DFT	DFT	L. Sevilla-Lara and E. Learned-Miller. Distribution Fields for Tracking. In CVPR, 2012. [www]
L1APG	L1APG	C. Bao, Y. Wu, H. Ling, and H. Ji. Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach. In CVPR, 2012. [www]
LOT	LOT	S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan. Locally Orderless Tracking. In CVPR, 2012. [www]
MTT	MTT	T.Zhang, B. Ghanem,S. Liu,and N. Ahuja. Robust Visual Tracking via Multi-task Sparse Learning. In CVPR, 2012. [www]
ORIA	ORIA	Y. Wu, B. Shen, and H. Ling. Online Robust Image Alignment via Iterative Convex Optimization. In CVPR, 2012. [www]
SCM	SCM	W. Zhong, H. Lu, and M.-H. Yang. Robust Object Tracking via Sparsity-based Collaborative Model. In CVPR, 2012. [www]
CSK	CSK	F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In ECCV, 2012. [www]
CT	CT	K. Zhang, L. Zhang, and M.-H. Yang. Real-time Compressive Tracking. In ECCV, 2012. [www]

Evaluated Trackers

Method	Representation	Search	MU	Code	FPS
CPF [44]	L, IH	PF	N	C	109
LOT [43]	L, color	PF	Y	M	0.70
IVT [47]	H, PCA, GM	PF	Y	MC	33.4
ASLA [30]	L, SR, GM	PF	Y	MC	8.5
SCM [65]	L, SR, GM+DM	PF	Y	MC	0.51
L1APG [10]	H, SR, GM	PF	Y	MC	2.0
MTT [64]	H, SR, GM	PF	Y	M	1.0
VTD [33]	H, SPCA, GM	MCMC	Y	MC-E	5.7
VTS [34]	L, SPCA, GM	MCMC	Y	MC-E	5.7
LSK [36]	L, SR, GM	LOS	Y	M-E	5.5
ORIA [58]	H, T, GM	LOS	Y	M	9.0
DFT [49]	L, T	LOS	Y	M	13.2
KMS [16]	H, IH	LOS	N	C	3,159
SMS [14]	H, IH	LOS	N	C	19.2
VR-V [15]	H, color	LOS	Y	MC	109
Frag [1]	L, IH	DS	N	C	6.3
OAB [22]	H, Haar, DM	DS	Y	C	22.4
SemiT [23]	H, Haar, DM	DS	Y	C	11.2
BSBT [50]	H, Haar, DM	DS	Y	C	7.0
MIL [5]	H, Haar, DM	DS	Y	C	38.1
CT [63]	H, Haar, DM	DS	Y	MC	64.4
TLD [31]	L, BP, DM	DS	Y	MC	28.1
Struck [26]	H, Haar, DM	DS	Y	C	20.2
CSK [27]	H, T, DM	DS	Y	M	362
CXT [18]	H, BP, DM	DS	Y	C	15.3

Benchmark dataset

- 50 sequences
  - 29491 frames



Benchmark dataset

- 11 attributes are defined for tracking sequence
  - illumination variations
  - scale variations
  - occlusions
  - deformation
  - motion blur
  - fast motion
  - in-plane rotation
  - out-of-plane rotation
  - out-of-view
  - background clutters
  - low resolution

Attr	Description
IV	Illumination Variation - the illumination in the target region is significantly changed.
SV	Scale Variation - the ratio of the bounding boxes of the first frame and the current frame is out of the range $[1/t_s, t_s]$, $t_s > 1$ (t_s=2).
OCC	Occlusion - the target is partially or fully occluded.
DEF	Deformation - non-rigid object deformation.
MB	Motion Blur - the target region is blurred due to the motion of target or camera.
FM	Fast Motion - the motion of the ground truth is larger than t_m pixels (t_m=20).
IPR	In-Plane Rotation - the target rotates in the image plane.
OPR	Out-of-Plane Rotation - the target rotates out of the image plane.
OV	Out-of-View - some portion of the target leaves the view.
BC	Background Clutters - the background near the target has the similar color or texture as the target.
LR	Low Resolution - the number of pixels inside the ground-truth bounding box is less than t_r (t_r=400).

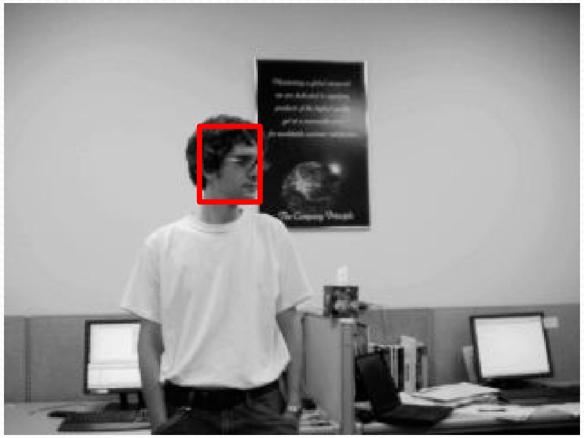
Attributes

- Illumination variations
  - Compared with the first frame, the illumination in the target region is changed visually

illumination variations

Attributes

- Out-of-plane rotation
  - Compared with the first frame, the target rotates out of the image plane



out-of-plane rotation

Attributes

- In-plane rotation
  - Compared with the first frame, the target rotates in the image plane

in-plane rotation

Attributes

- Deformation
  - non-rigid deformation occurs in the frame

deformation

Attributes

- Scale variations
  - The ratio of the number of pixels in the bounding box of 1st frame to current frame is not less than a threshold t or not larger than 1/t (e.g. t=2)

scale variations

Attributes

- Out-of-view
  - One portion of the target is out of the image region
    - At present, we only consider partial out-of-view. Our dataset does not include the sequence where the target is totally out-of view
    - The annotation bounding box is inside the image plane

out-of-view

Attributes

- Background clutters
  - The background near the target has the similar color or texture as the target

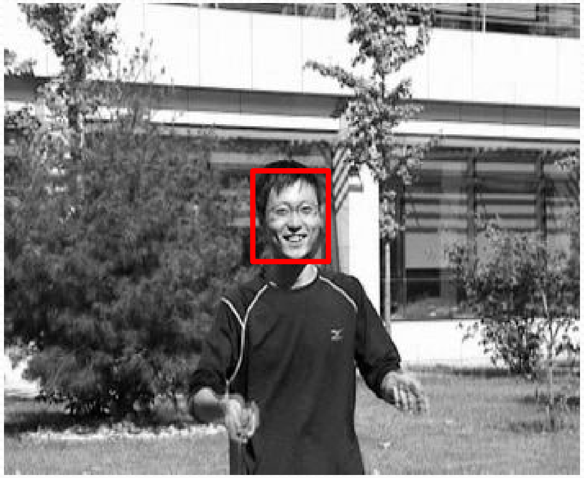
background clutter

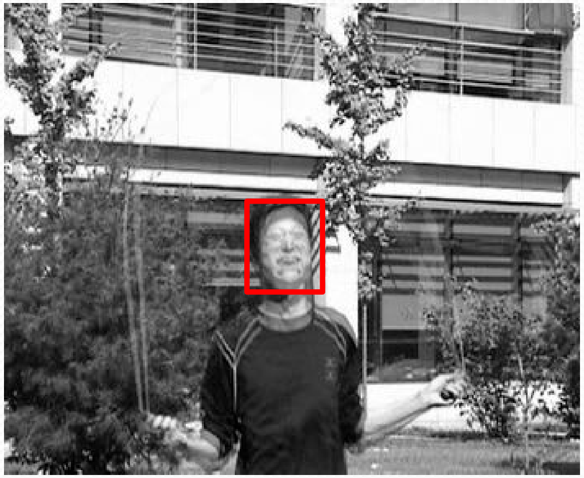
Attributes

- Low resolution
  - The number of pixels inside the groundtruth bounding box is less than a threshold (e.g. 400)

Attributes

- Motion blur
  - The target region is blurred due to the motion of target or camera
- Fast motion
  - The motion computed from the ground truth is larger than t pixels (e.g. t=20)

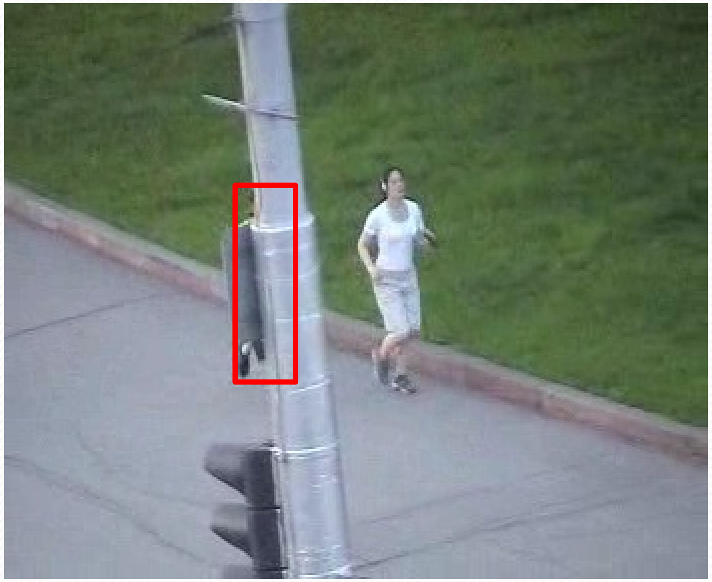




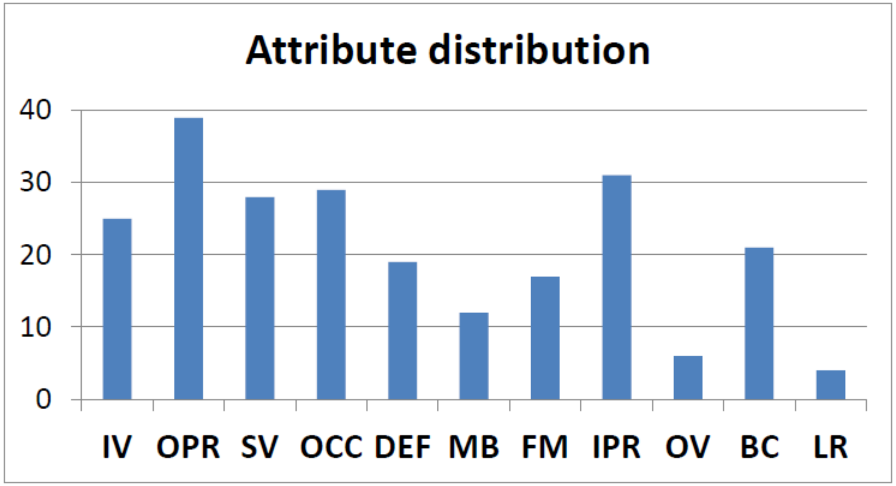
blur & fast motion

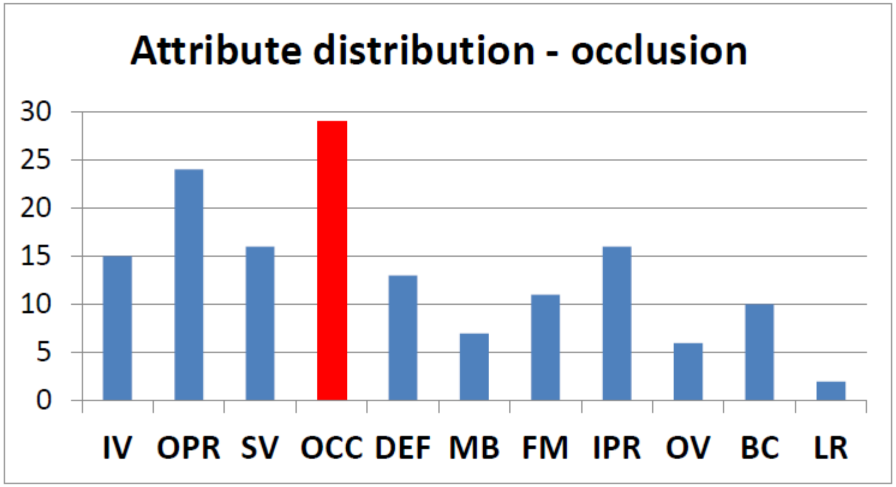
Attributes

- Occlusions
  - The target is occluded by other objects
  - Even though the target is fully occluded in one frame, we also annotate the bounding box



Attribute Distribution





Evaluation Methodology

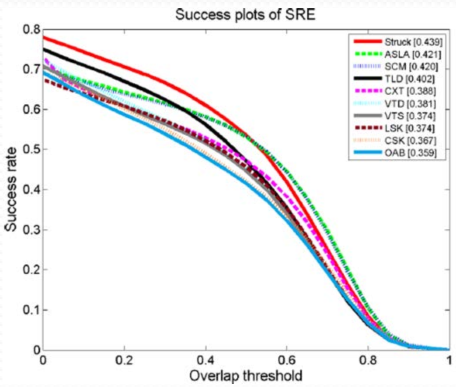
- Metrics
  - Bounding box overlap (success plot)

$$S = \frac{| r_t \bigcap r_0 |}{| r_t \bigcup r_0 |}$$

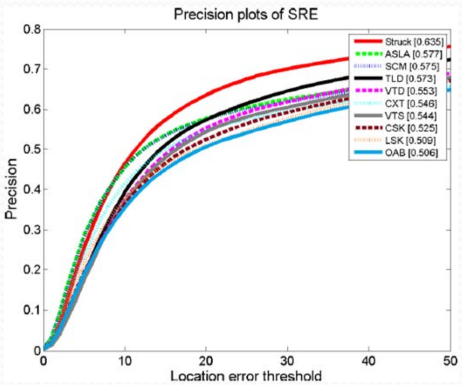
r_t : a tracked bounding box

r_0 : the ground-truth bounding box

  - Center location error (precision plot)
    - Euclidian distance between the center of tracking result and the center of annotation



success plot



precision plot

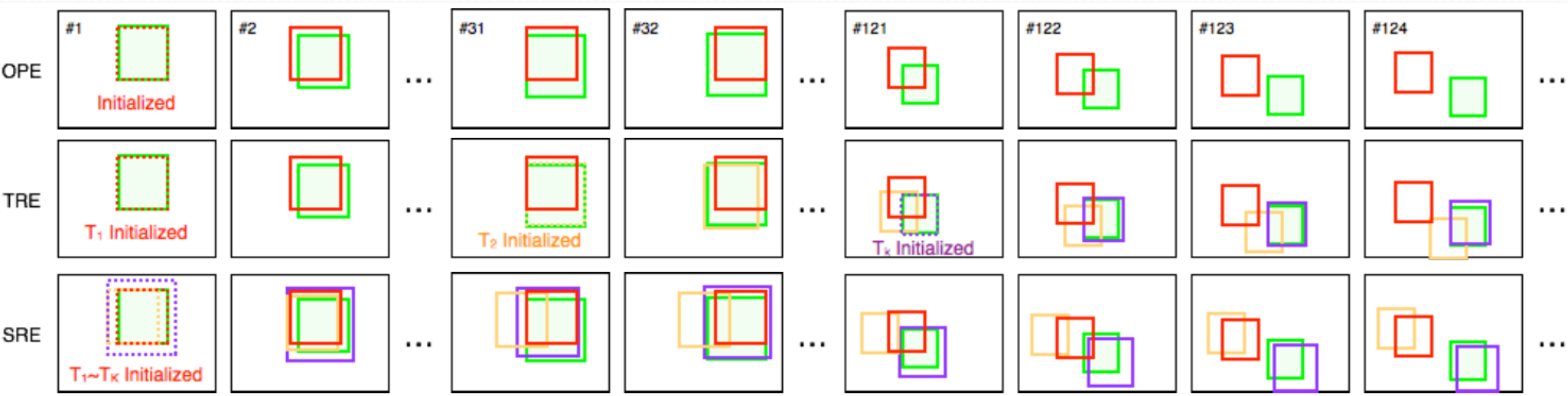
Evaluation Methodology

- One-Pass Evaluation (OPE)
  - The conventional way to evaluate trackers
  - Only one initialization
  - Bias or randomness
- **Robustness Evaluation**
  - Temporal Robustness Evaluation (TRE):
    - Sampling the frames for tracking initialization on each sequence
  - Spatial Robustness Evaluation (SRE):
    - Sampling the initial bounding box in the first frame by shifting or scaling the ground truth

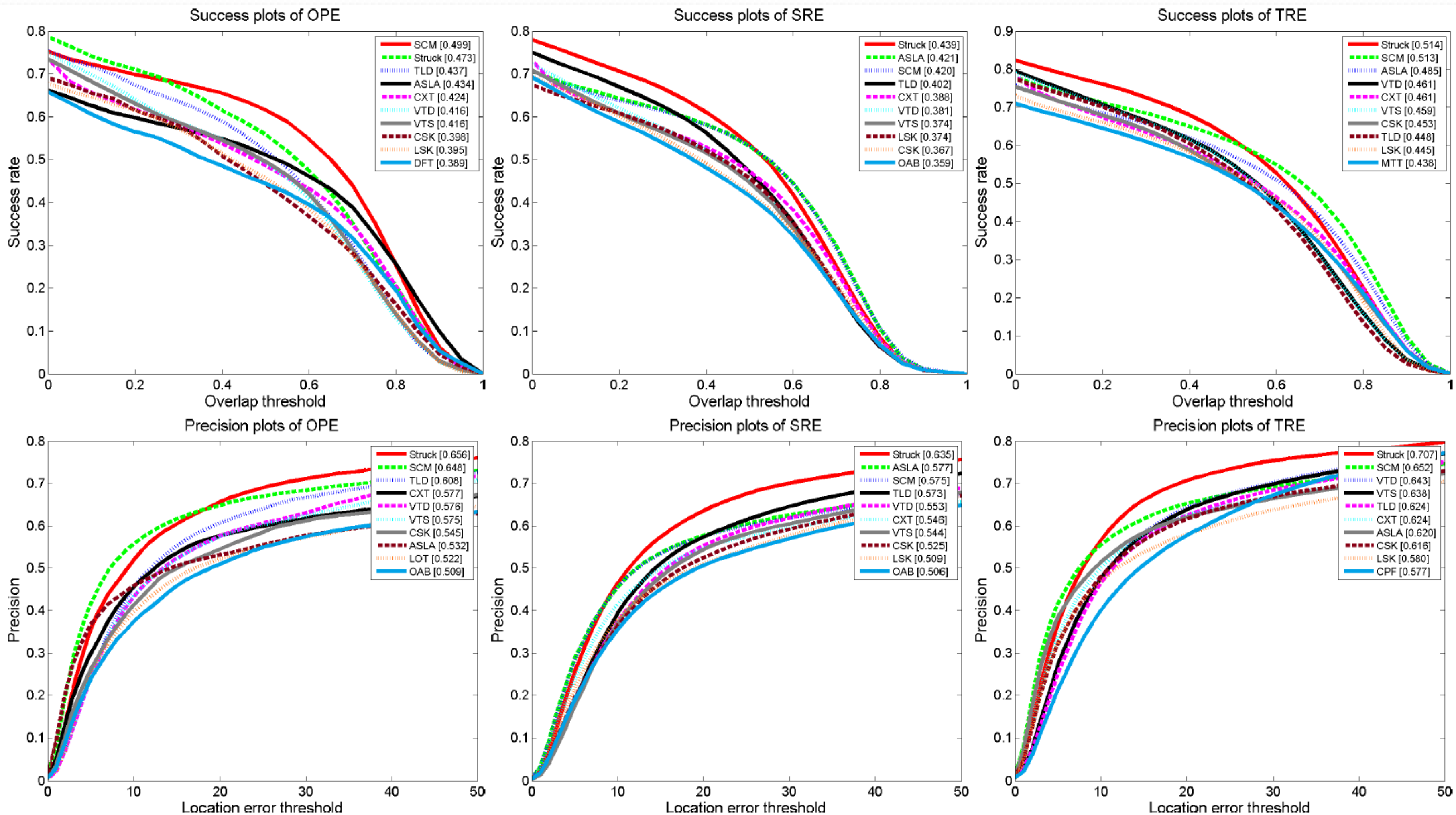
Sensitivity to the initialization

Evaluation Methodology

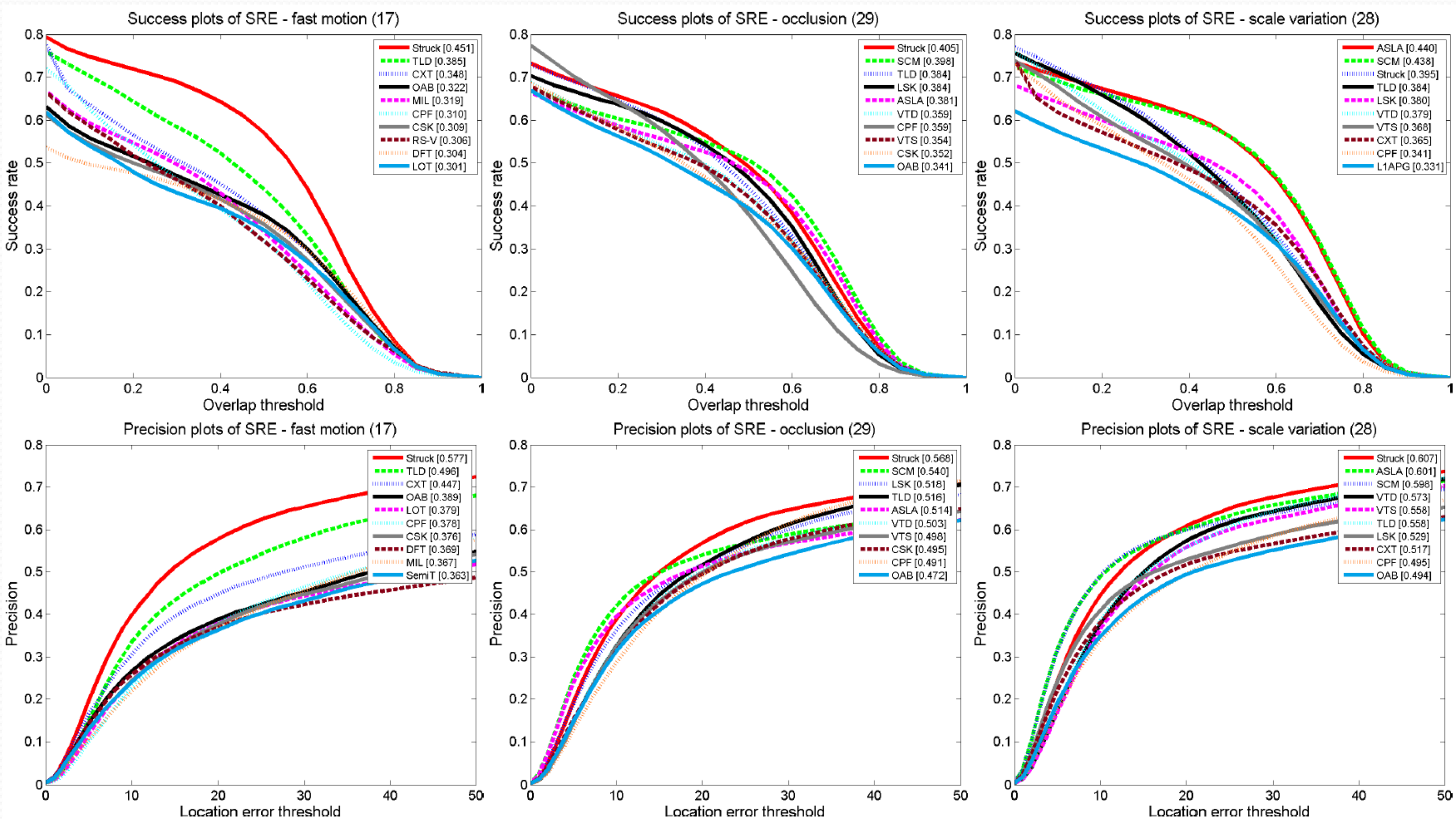
- One-Pass Evaluation (OPE)
- **Robustness Evaluation**
  - Temporal Robustness Evaluation (TRE)
  - Spatial Robustness Evaluation (SRE)



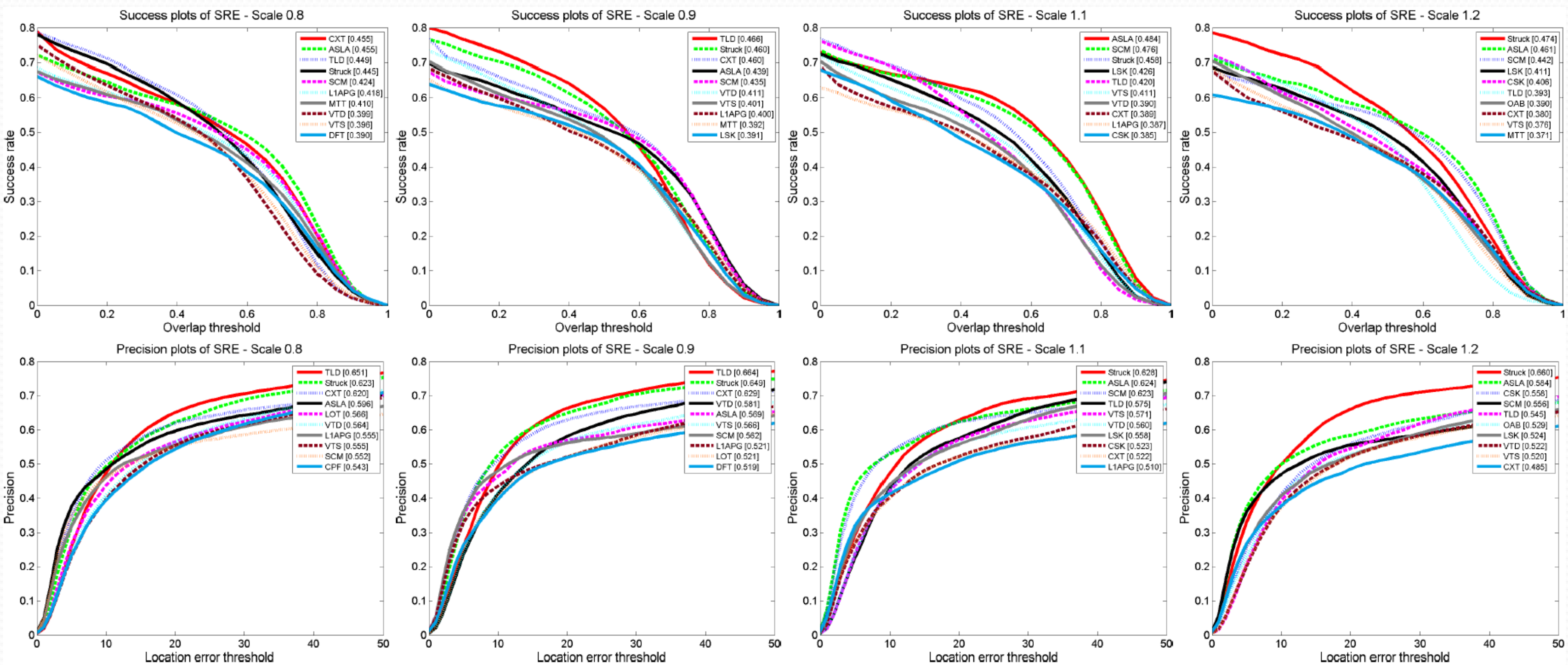
Evaluation Results



Attribute-based Performance Analysis

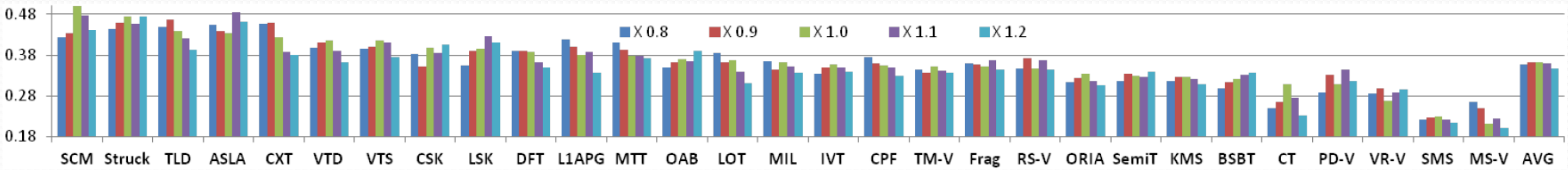


Initialization with Different Scales



Initialization with Different Scale

- Performance decreases with the increase of initialization scale
  - TLD, CXT, DFT and LOT
- Some perform better when the scale factor is smaller
  - L1APG, MTT, LOT and CPF
- Some trackers perform well or even better when the initial bounding box is enlarged
  - Struck, OAB, SemiT, and BSBT



Concluding Remarks

- Some tracking components that are essential for improving tracking performance.
  - Background information: serving as discriminative model or context
  - Local models: effective for handling partial occlusion or deformation
  - Motion model or dynamic model: improving the tracking efficiency
- Tracking evaluation platform http://visual-tracking.net/
  - Tracker library
  - Dataset with annotation

Concluding Remarks

- Tracking evaluation platform
  - Tracker library
  - Dataset with annotation
  - Evaluation toolkit

http://visual-tracking.net/

Q&A