Monocular Multiview Object Tracking with 3D Aspect Parts

Yu Xiang and Silvio Savarese

Computational Vision and Geometry Lab

Stanford University

1st Stanford-SNU Workshop on Automated Driving

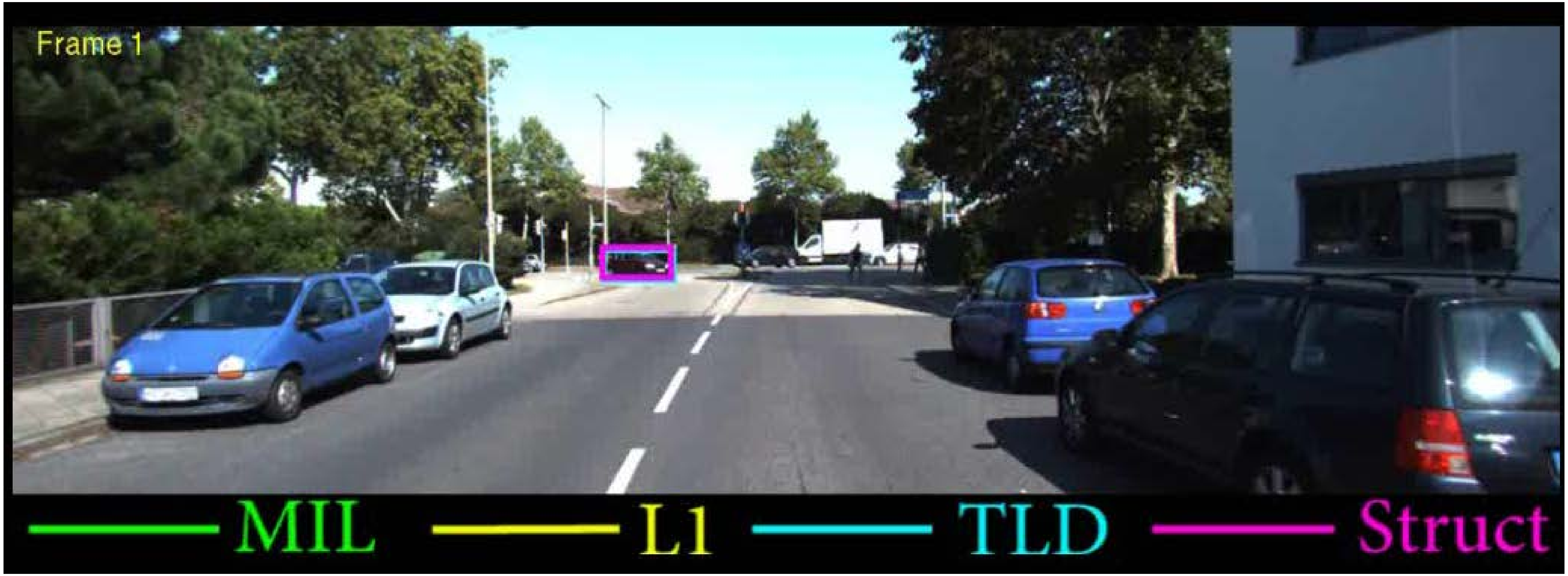
Car Tracking in Autonomous Driving

Cars change their viewpoints/poses!

How to robustly track the location and 3D pose of a car?

How to identify functional portions of the object, such as a door or a window?

Online Object Tracking



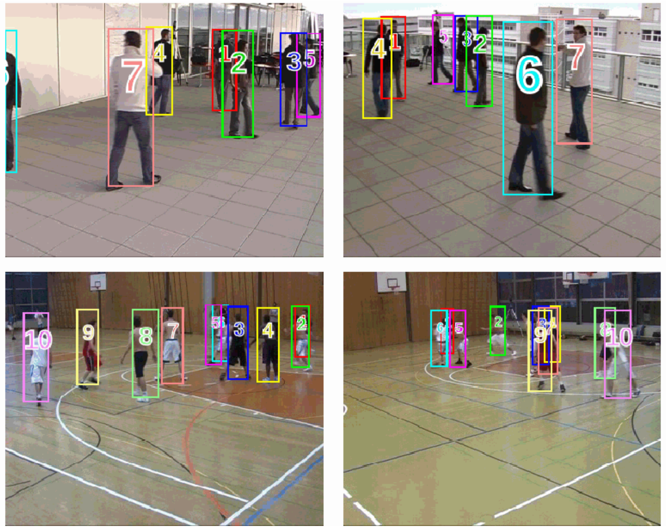
[MIL] Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. TPAMI, 2011.
[L1] Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In CVPR, 2012.
[TLD] Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. TPAMI, 2012.
[Struct] Hare, S., Saari, A., Torr, P.H.: Struck: Structured output tracking with kernels. In ICCV, 2011.

Ours: Monocular Multiview Object Tracking

Xiang, Y., Song, C., Mottaghi, R. and Savarese, S.: Monocular multiview object tracking with 3D aspect parts. In ECCV, 2014.

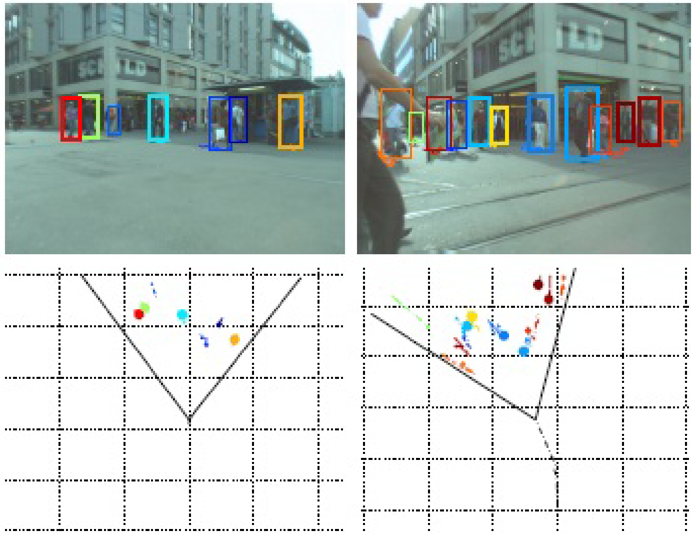
Related Work: Tracking by Detection

- Link detections from a category-level detector



K shortest paths
Berclaz et al., TPAMI'11

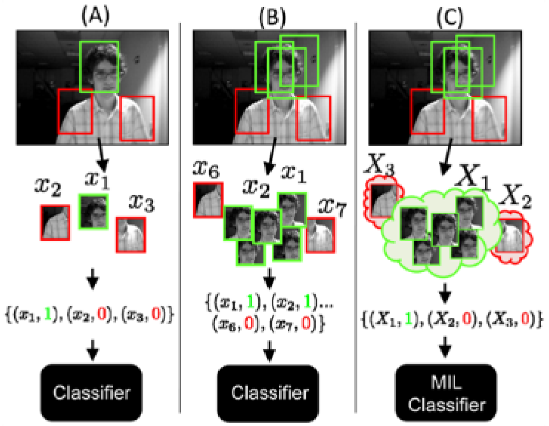
Continuous energy minimization
Andriyenko and Schindler, CVPR'11



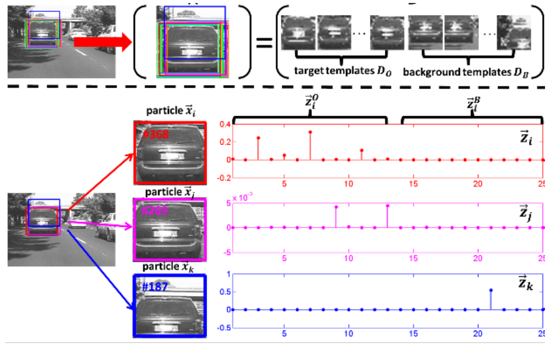
RJMCMC particle filtering
Choi et al., TPAMI'13

Related Work: Online Object Tracking

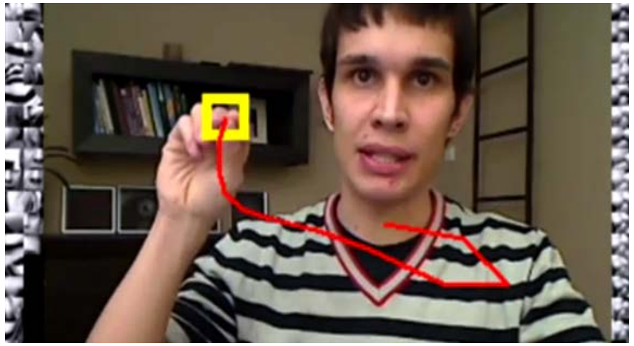
- Learn object appearance model online


(A)
(B)
(C)
x_2 x_1 x_3
x_6 x_2 x_1 x_7
X_3 X_1 X_2
$\{(x_1,1),(x_2,0),(x_3,0)\}$
$\{(x_1,1),(x_2,1)\ldots(x_6,0),(x_7,0)\}$
$\{(X_1,1),(X_2,0),(X_3,0)\}$
Classifier
Classifier
MIL Classifier

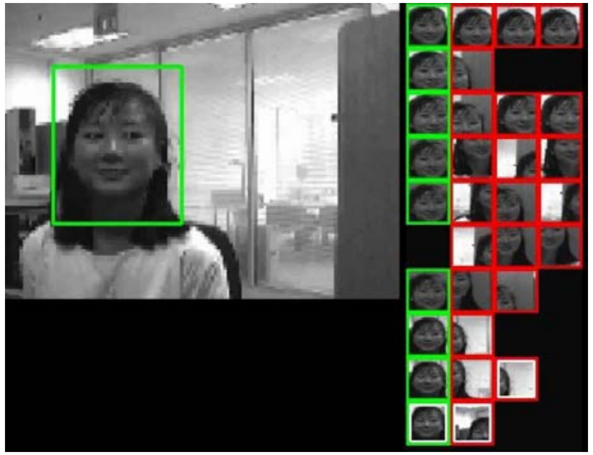
Multiple instance learning
Babenko et al., TPAMI’11.



L1 Tracker
Ling et al., ICCV’09, CVPR’12

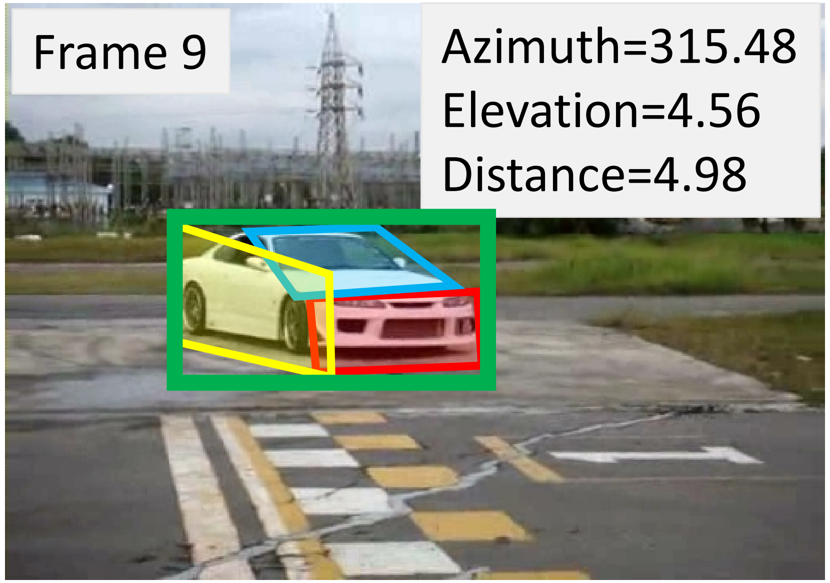


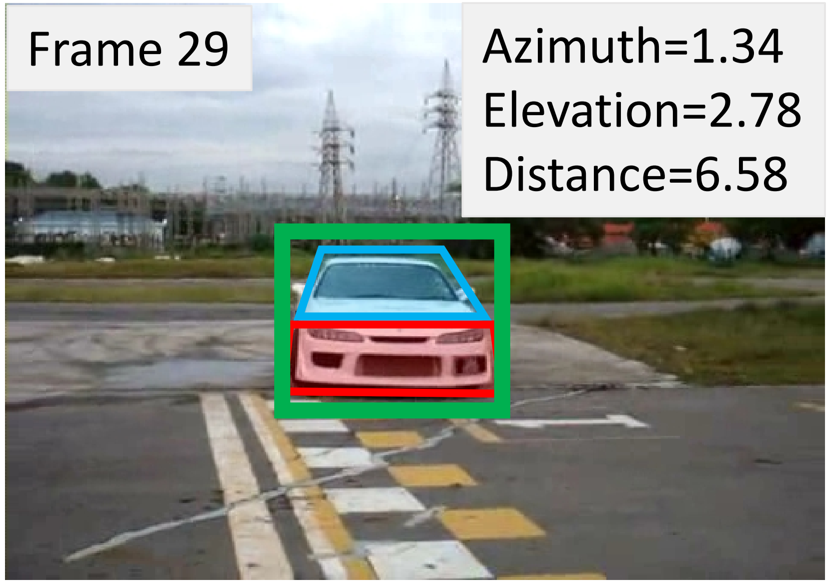
TLD Tracker
Kalal et al., TPAMI’12

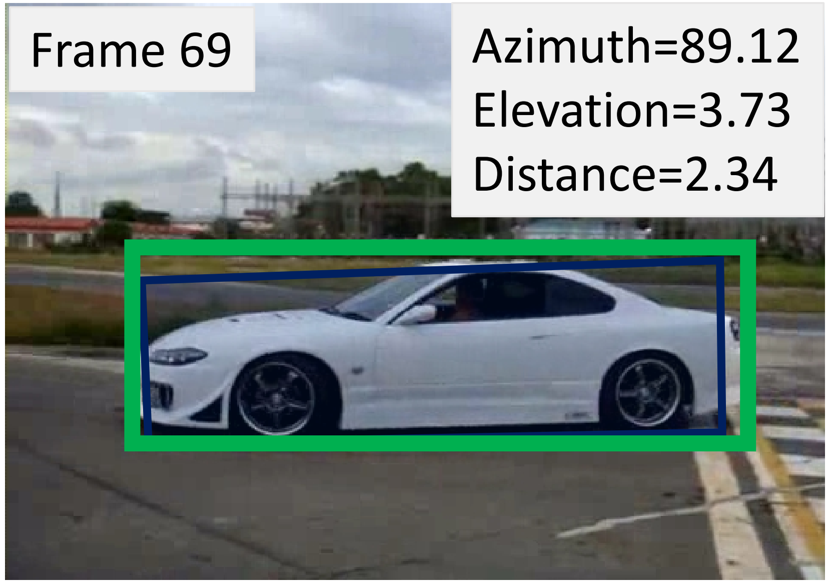


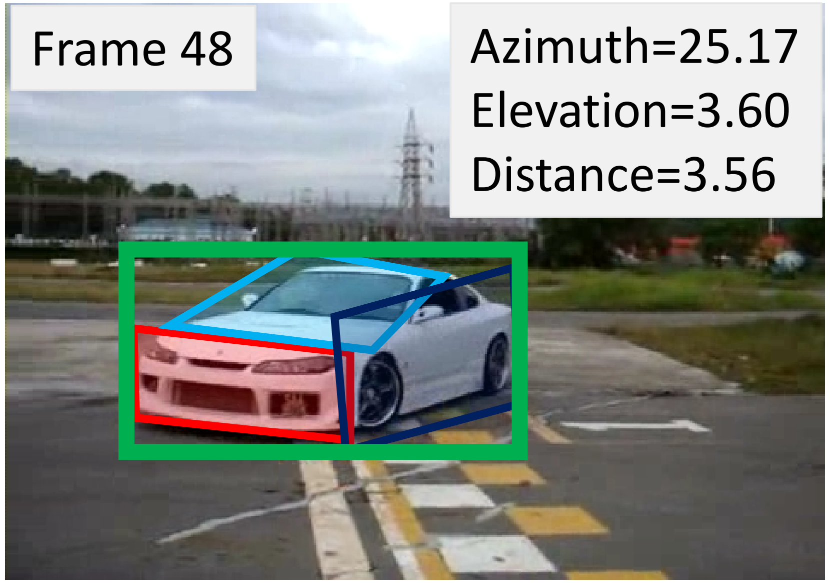
Struct Tracker
Hare et al., ICCV’11

Goal: Track viewpoint and parts of the target

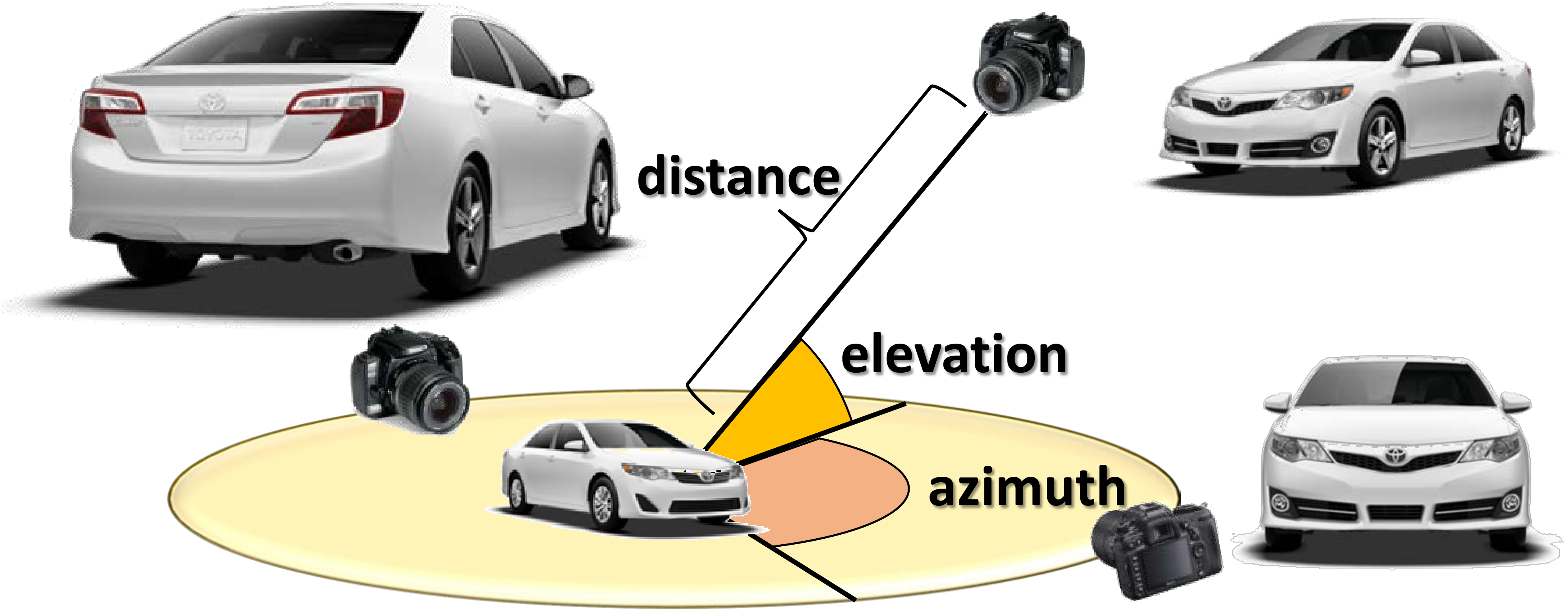




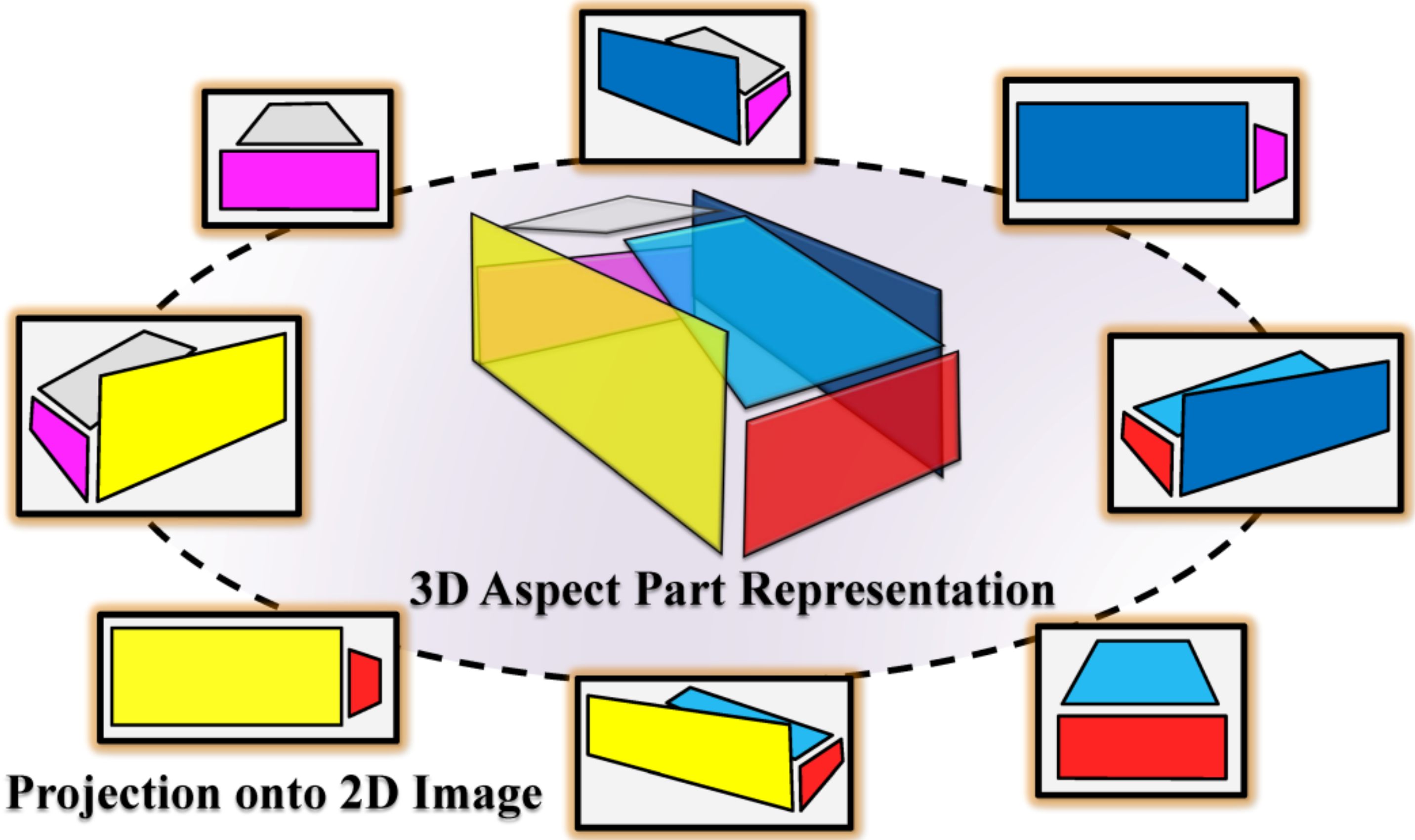




Viewpoint Representation



Object Representation



Xiang, Y. and Savarese, S.: Estimating the aspect layout of object categories. In CVPR, 2012.

Multiview Tracking Framework

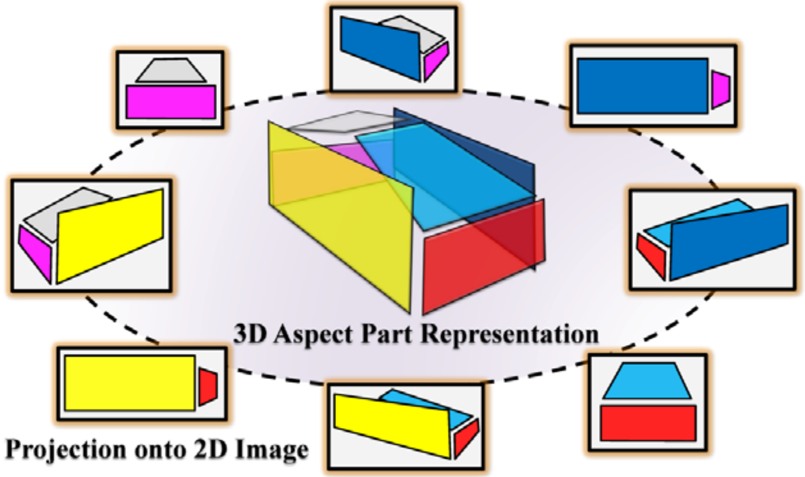
- Posterior distribution (recursive Bayesian filtering)

$$P(\overset{\text{Part locations}}{X_t}, \overset{\text{Viewpoint}}{V_t} \mid \overset{\text{Video frames}}{Z_{1:t}})$$

$$\propto \underbrace{P(Z_t \mid X_t, V_t)}_{\text{Likelihood}} \int \underbrace{P(X_t, V_t \mid X_{t-1}, V_{t-1})}_{\text{Motion prior}} \underbrace{P(X_{t-1}, V_{t-1} \mid Z_{1:t-1})}_{\text{Posterior at t-1}} dX_{t-1} dV_{t-1}$$

Multiview Tracking Framework

- Likelihood



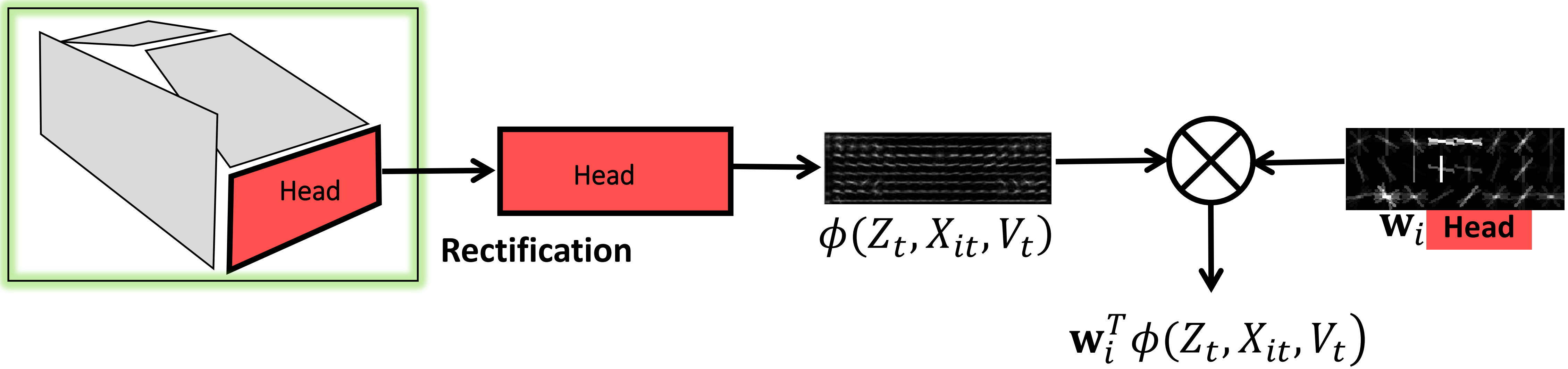
$$P(Z_t \mid X_t, V_t) = \prod_{i=1}^{n} P(Z_t \mid X_{it}, V_t)$$

$$P(Z_t \mid X_{it}, V_t) \propto \exp\left(\Lambda_{\text{category}}(Z_t, X_{it}, V_t) + \Lambda_{\text{online}}(Z_t, X_{it}, V_t)\right)$$

Multiview Tracking Framework

- Category-level part templates

$$\Lambda_{\text{category}}(Z_t, X_{it}, V_t) = \begin{cases} \mathbf{w}_i^{\mathrm{T}} \phi(Z_t, X_{it}, V_t), & \text{if visible} \\ \alpha_i, & \text{if self-occluded} \end{cases}$$



Head

Head

Rectification

$\phi(Z_t, X_{it}, V_t)$

$\mathbf{w}_i$ Head

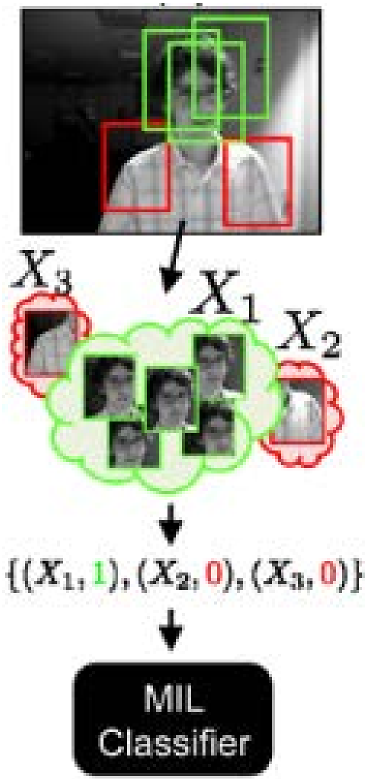
$\mathbf{w}_i^T \phi(Z_t, X_{it}, V_t)$

Multiview Tracking Framework

- Online-learned appearance model

$$\Lambda_{\text{online}}(Z_t, X_{it}, V_t) = \begin{cases} \mathbf{H}_i(\psi(Z_t, X_{it}, V_t)), \text{ if visible} \\ \lambda_0, \text{ if self-occluded} \end{cases}$$

- Multiple instance learning classifier [1]



[1] Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. TPAMI, 2011.

Multiview Tracking Framework

- Motion prior

$$P(X_t, V_t \mid X_{t-1}, V_{t-1}) = \underbrace{P(X_t \mid X_{t-1}, V_t)}_{\text{Location Motion}} \; \underbrace{P(V_t \mid V_{t-1})}_{\text{Viewpoint Motion}}$$

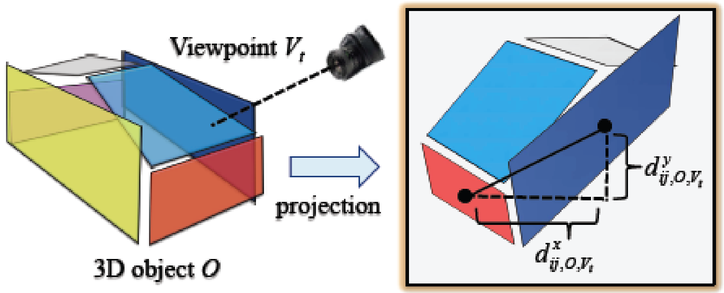
MRF

$$P(X_t \mid X_{t-1}, V_t) \propto \prod_{i=1}^{n} P(X_{it} \mid X_{i(t-1)}) \prod_{(i,j)} \Lambda(X_{it}, X_{jt}, V_t)$$

$$P(X_{it} \mid X_{i(t-1)}) \sim N(X_{i(t-1)}, \sigma_x^2, \sigma_y^2)$$

Pair-wise

$$P(V_t \mid V_{t-1}) \sim N(V_{t-1}, \sigma_a^2, \sigma_e^2, \sigma_d^2)$$



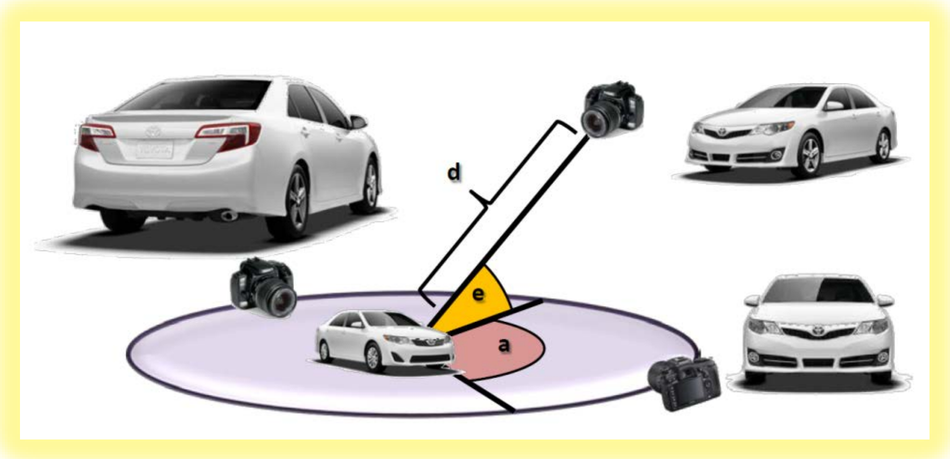
$$\Lambda(X_{it}, X_{jt}, V_t) = P(\Delta_t(x_i, x_j) \mid V_t) P(\Delta_t(y_i, y_j) \mid V_t)$$

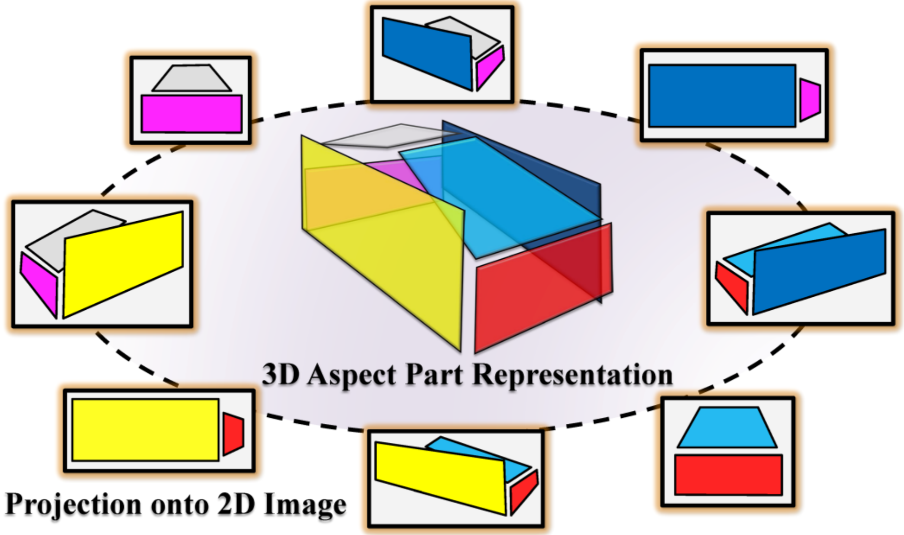
$$P(\Delta_t(x_i, x_j) \mid V_t) \sim N(d^x_{ij,O,V_t}, \sigma_{dx}^2)$$

$$P(\Delta_t(y_i, y_j) \mid V_t) \sim N(d^y_{ij,O,V_t}, \sigma_{dy}^2)$$

Multiview Tracking Framework

- Particle filtering tracking
  - ❑MCMC sampling
  - ❑Sample viewpoint
  - ❑Check part visibility
  - ❑Sample part locations





Experiments

- Datasets

□ A new YouTube dataset (9 sequences)

□ Subset of KITTI [1] (11 sequences)

[1] Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.

2D Object Tracking

Online tracking Object detection + particle filtering

Video	MIL [1]	L1 [2]	TLD [3]	Struct [4]	DPM [5]+PF	Ours w/o online	Ours with online
YouTube	0.37	0.44	0.38	0.40	0.74	0.74	**0.75**
KITTI [6]	0.34	0.28	0.29	0.36	0.54	0.55	**0.58**

Metric: mean bounding box overlap ratio

[1] Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. TPAMI, 2011.
[2] Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In CVPR, 2012.
[3] Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. TPAMI, 2012.
[4] Hare, S., Saari, A., Torr, P.H.: Struck: Structured output tracking with kernels. In ICCV, 2011.
[5] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. TPAMI, 2010.
[6] Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.

Viewpoint and 3D Aspect Part

❑ Viewpoint estimation error

Video	Ours with online	Ours w/o online	ALM [1]
YouTube	**13.46°**	18.38°	47.24°
KITTI	**14.66°**	23.20°	37.89°

Metric: mean absolute difference in azimuth angle

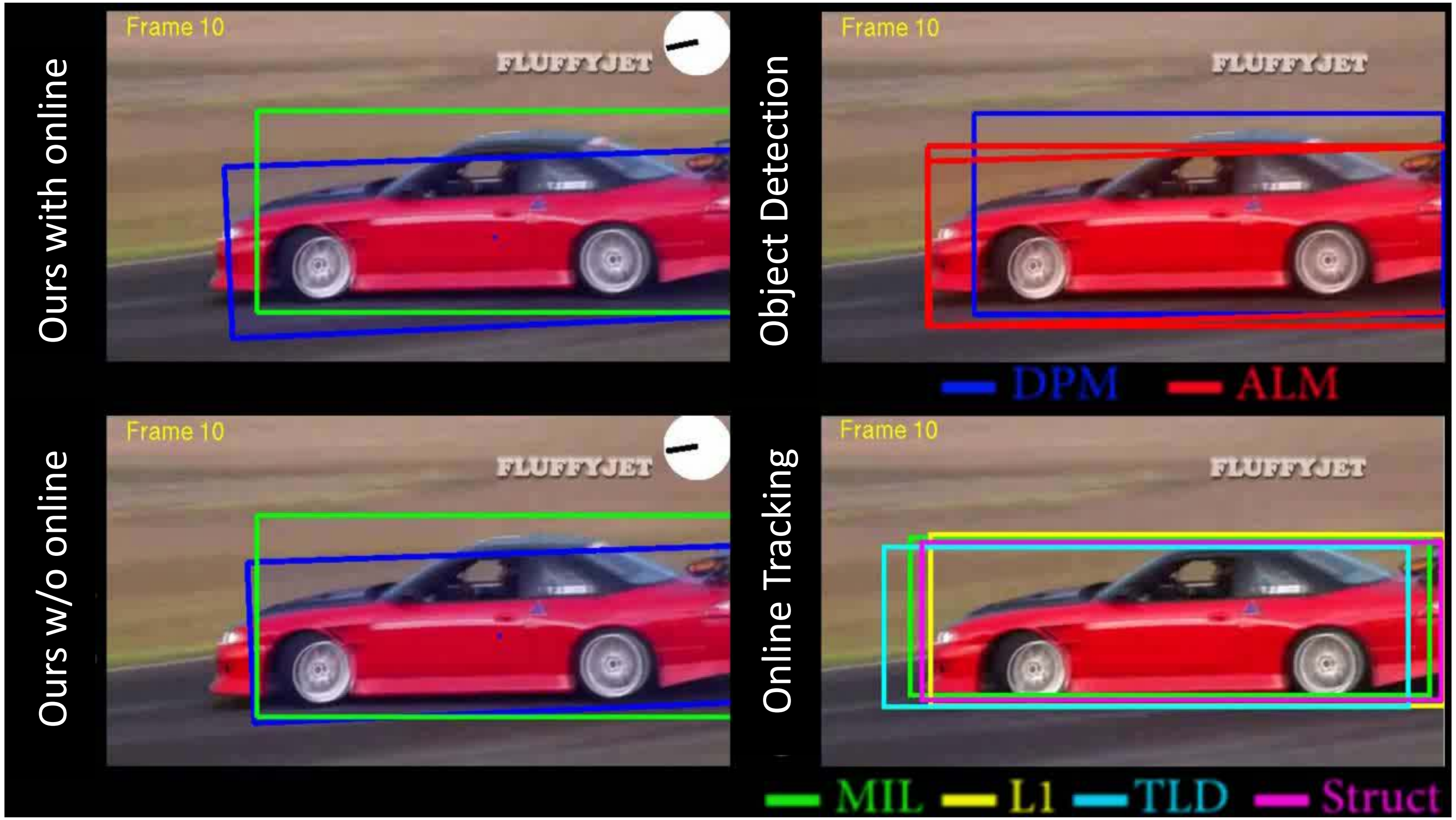
❑ 3D aspect part localization accuracy

Video	Ours with online	Ours w/o online	ALM [1]
YouTube	**0.41**	0.40	0.30
KITTI	**0.36**	0.30	0.26

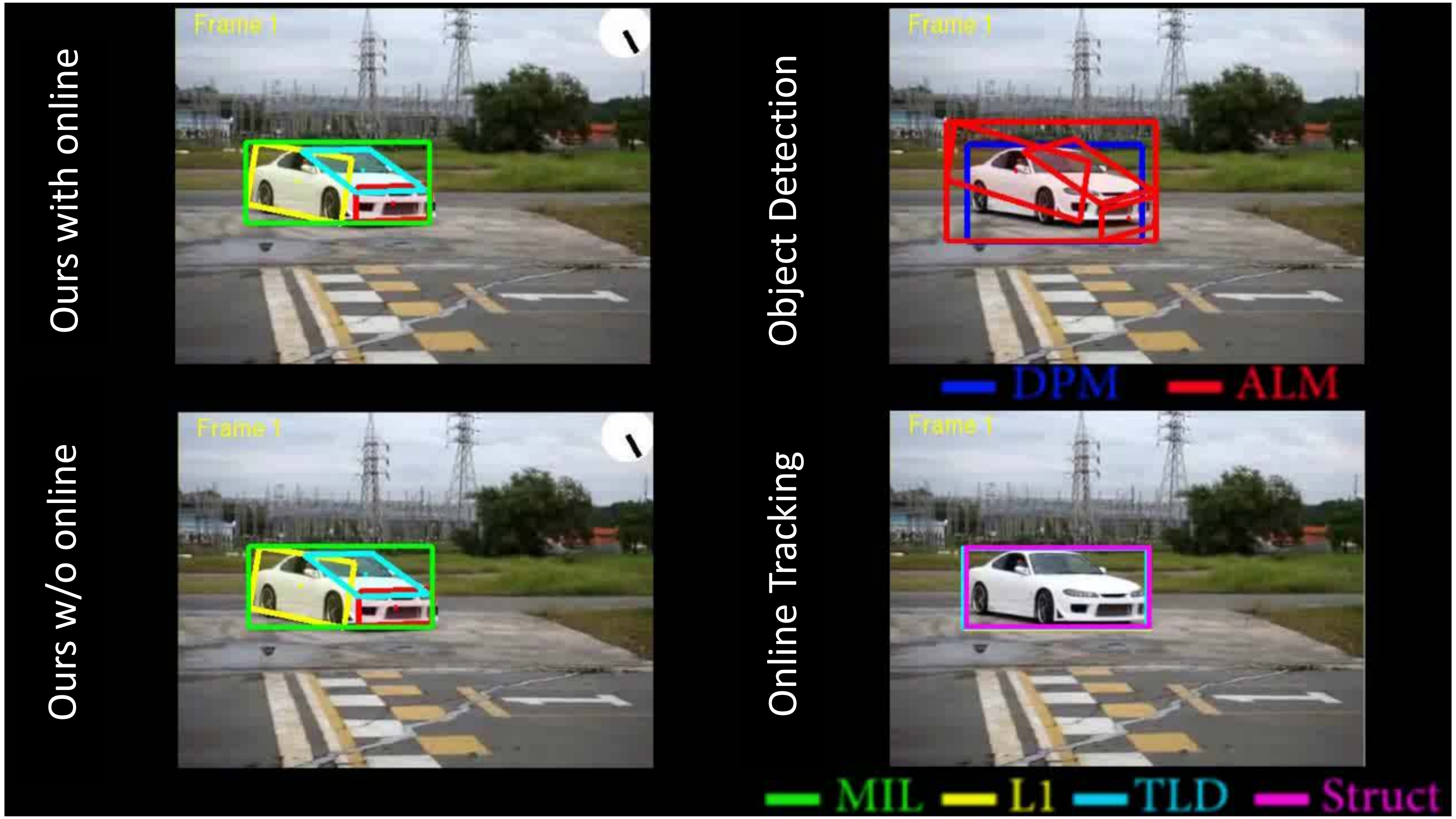
Metric: mean overlap ratio of part shape

[1] Xiang, Y., Savarese, S.: Estimating the aspect layout of object categories. In CVPR, 2012.

Result Videos



Result Videos



Result Videos

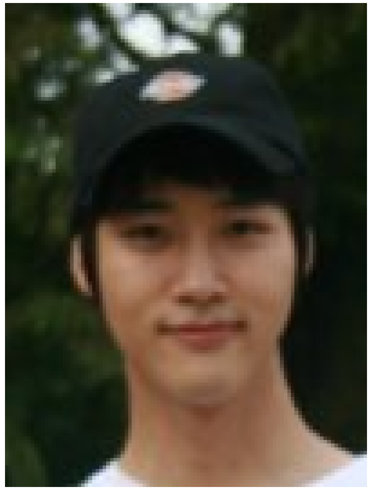
Conclusion

- Propose a new multiview object tracking framework

- Track viewpoint and 3D aspect parts in time

- Apply to vehicle tracking in autonomous driving scenarios

Acknowledgements



Changkyu Song

Roozbeh Mottaghi

CAREER grant N.1054127

Thank you!