# Tracking Revisited using RGBD Camera: Unified Benchmark and Baselines

PRINCETON VISION GROUP

Shuran Song Jianxiong Xiao

## Motivation

### 1.Traking is difficult but depth may help

- Popularity of depth sensors made it possible to obtain reliable depth easily
- Game changer for tracking?
- Prevent model drift & handle occlusion?

### 2. A Good benchmark is important

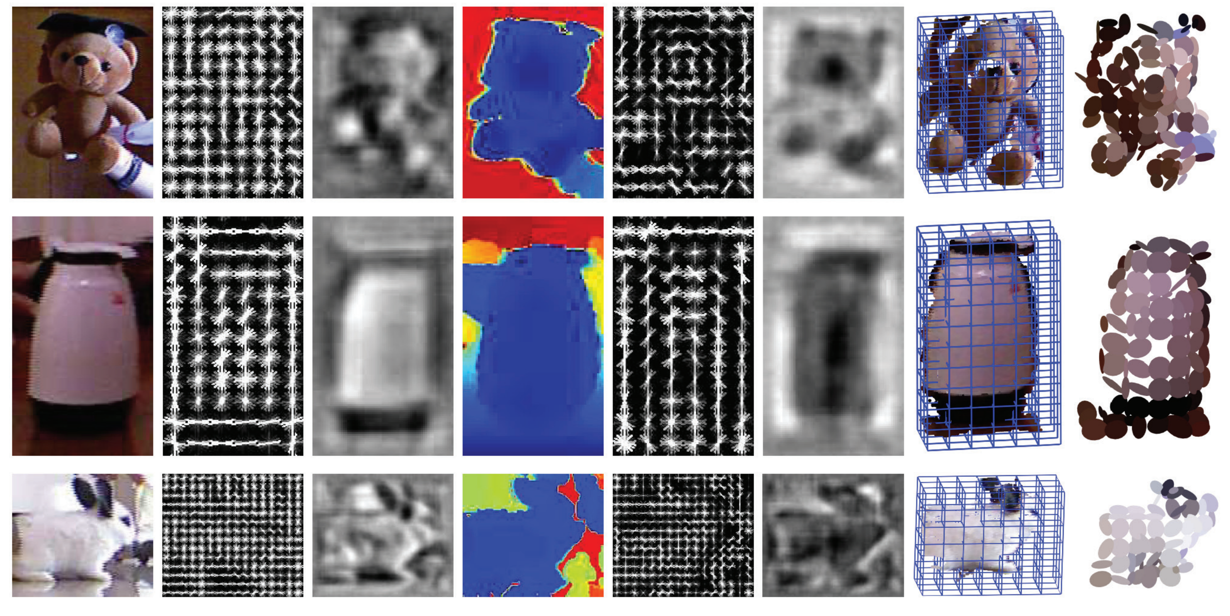
Previous benchmarks
- Small
- Inconsistently labeled
- No depth

### Contribution

- We construct a unified benchmark dataset of 100 RGBD videos with high diversity
- We propose various kinds of RGBD tracking algorithms in both 2D and 3D, and present a quantitative comparison

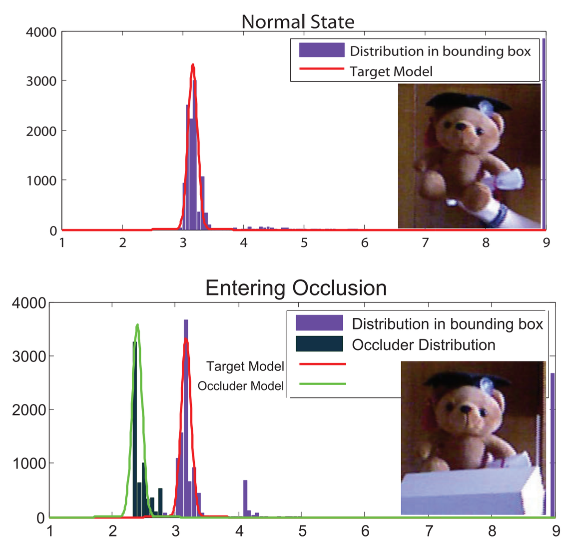
**Data + Code:** http://tracking.cs.princeton.edu

## Features



RGB → HOG → iHOG Depth → HOG → iHOG Point Cloud → 3D

## Occlusion Handling




The depth histogram of all pixels inside a bounding box can be approximated as a Gaussian distribution:

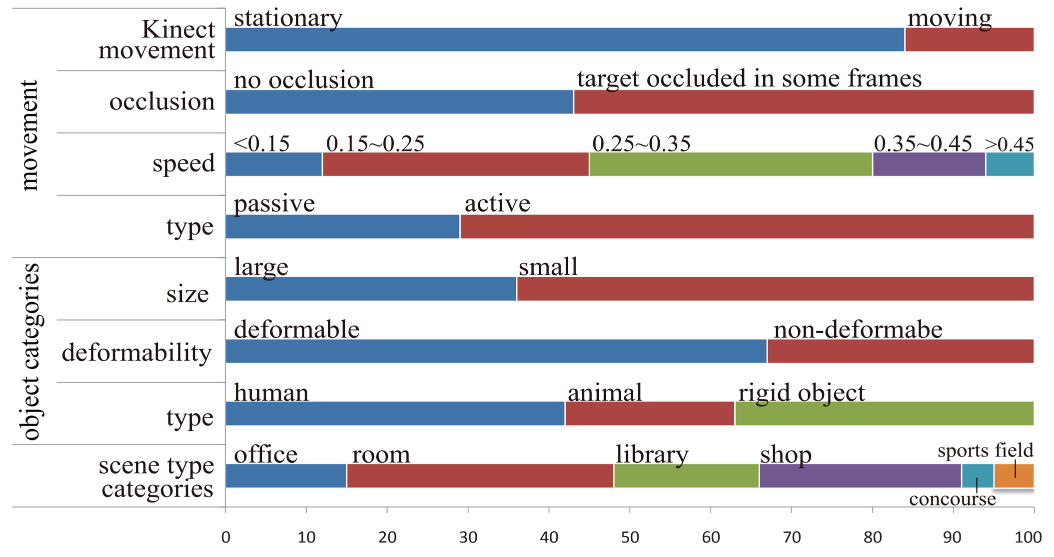
$$h_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

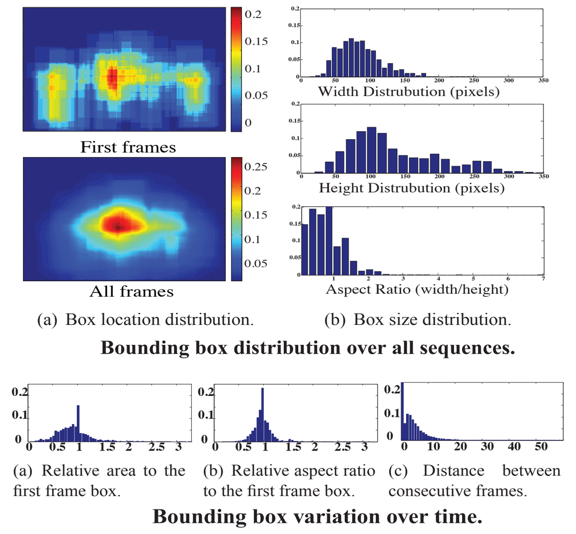
The likelihood of occlusion:

$$O_i = \sum_{d=0}^{\mu_i - \sigma_i} h_i(d) / \sum_d h_i(d)$$

Red Gaussians: the target model
Green Gaussian: the occluder model

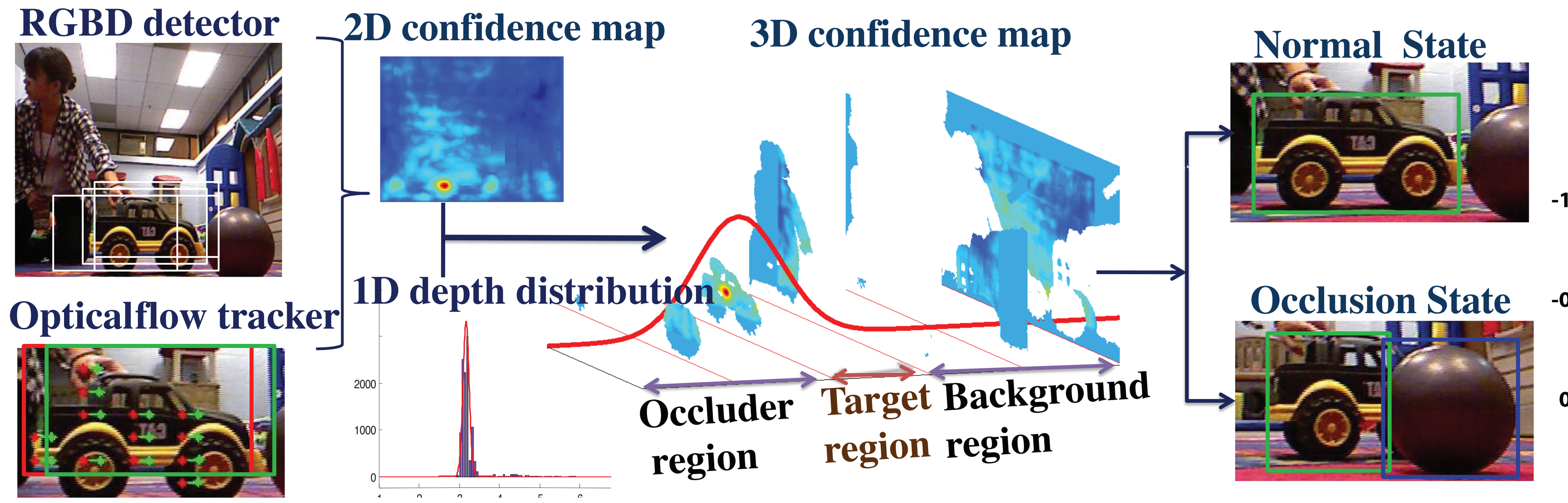
## Princeton Tracking Benchmark



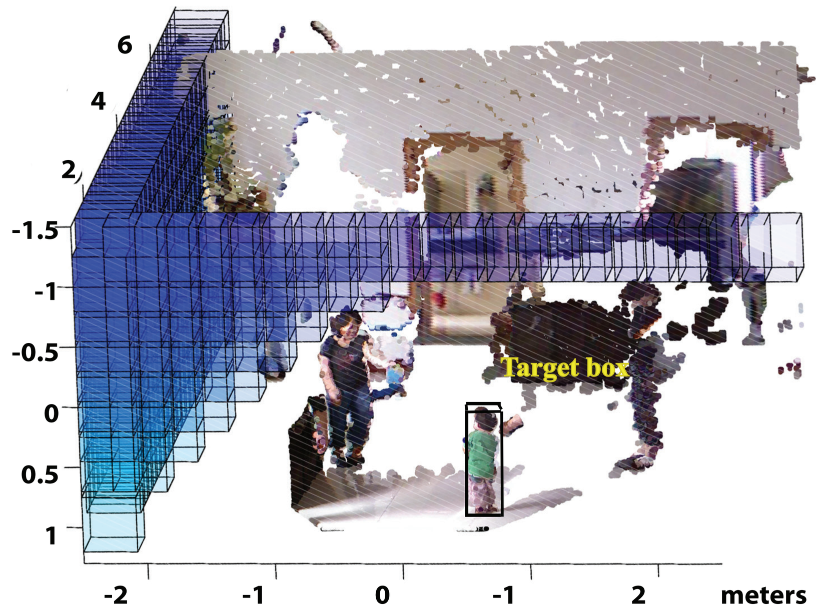




100 RGBD videos, including deformable objects, various occlusion conditions, moving camera, and different scenes.

## Baseline Algorithms




RGBD tracking algorithm based on 2D image patch




3D sliding window detection

## Evaluation

### Successful Rate

$$r_i = \begin{cases} \frac{\text{area}(ROI_{T_i} \cap ROI_{G_i})}{\text{area}(ROI_{T_i} \cup ROI_{G_i})} & \text{if both } ROI_{T_i} \text{ and } ROI_{G_i} \text{ exist} \\ 1 & \text{if neither } ROI_{T_i} \text{ and } ROI_{G_i} \text{ exist} \\ -1 & \text{otherwise} \end{cases}$$

$ROI_{T_i}$ : tracking bounding box,
$ROI_{G_i}$ is the ground truth bounding box.
By setting a minimum overlapping area $r_t$, we can calculate the average success rate $R$:

$$R = \frac{1}{N}\sum_{i=1}^{N} u_i, \quad \text{where} \quad u_i = \begin{cases} 1 & \text{if } r_i > r_t \\ 0 & \text{otherwise} \end{cases},$$

### Error Type

Type I : $ROI_{T_i} \neq null$ and $ROI_{G_i} \neq null$ and $r_i < r_t$
Type II : $ROI_{T_i} \neq null$ and $ROI_{G_i} = null$
Type III : $ROI_{T_i} = null$ and $ROI_{G_i} \neq null$

### RGB tracking Algorithms


[Struck] S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured output tracking with kernels. In ICCV, 2011
[VTD] J. Kwon and K. M. Lee. Visual tracking decomposition. In CVPR, 2010.
[CT] K. Zhang, L. Zhang, and M.-H. Yang. Real-time compressive tracking. In ECCV, 2012
[MIL] B. Babenko, M.-H. Yang, and S. Belongie. Visual Tracking with Online Multiple Instance Learning. In CVPR, 2009
[TLD] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. PAMI, 2012.
[SemiB] H. Grabner, C. Leistner, and H. Bischof. Semi-supervised on-line boosting for robust tracking. In ECCV, 2008


### Ranking

Evaluation results: successful rate (SR) % and corresponding rankings (in parentheses) under different categorizations.

| algorithm | avg. rank | all SR | target type | | | target size | | movement | | occlussion | | motion type | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | human | animal | rigid | large | small | slow | fast | yes | no | active | passive |
| GTbestRatio | - | 99.665 | 99.9(1) | 100.0(1) | 99.3(1) | 99.8(1) | 99.6(1) | 99.5(1) | 99.7(1) | 99.6(1) | 99.7(2) | 99.7(1) | 99.5(1) |
| GTfirstRatio | - | 99.3 | 99.6(2) | 99.8(2) | 98.6(2) | 99.4(2) | 99.2(2) | 99.0(2) | 99.4(2) | 99.2(2) | 99.5(3) | 99.4(2) | 99.0(2) |
| GTbestSize | - | 97 | 95.0(3) | 99.3(3) | 98.1(3) | 96.2(3) | 97.6(3) | 98.4(3) | 96.5(3) | 96.4(3) | 97.8(4) | 96.3(3) | 98.8(3) |
| GTonOcc | - | 93.05 | 91.0(4) | 94.3(4) | 94.8(5) | 92.6(4) | 93.4(4) | 94.3(4) | 92.5(4) | 88.0(5) | 100.0(1) | 93.2(4) | 92.5(5) |
| GTfirstSize | - | 91.03 | 86.7(5) | 90.3(5) | 96.5(4) | 88.7(5) | 92.8(5) | 93.2(5) | 90.1(5) | 90.8(4) | 91.3(5) | 88.6(5) | 97.5(4) |
| RGBDOcc+OF | 1.18 | 73.3 | 74.0(1) | 62.6(2) | 78.4(1) | 78.1(1) | 69.7(1) | 76.3(1) | 72.2(1) | 72.0(1) | 75.2(2) | 70.0(1) | 82.3(1) |
| RGBD+OF | 1.91 | 68.1 | 63.9(2) | 65.3(1) | 74.5(2) | 71.5(2) | 65.5(2) | 73.4(2) | 65.9(2) | 60.1(2) | 79.0(1) | 65.8(2) | 74.0(3) |
| PC(det+flow) | 3.09 | 58.9 | 50.5(3) | 51.6(3) | 72.7(3) | 63.4(3) | 55.5(4) | 73.9(2) | 53.0(3) | 55.0(3) | 64.4(5) | 75.5(2) | 52.7(3) |
| RGB+OF | 4.82 | 53.2 | 47.1(4) | 47.0(6) | 63.6(4) | 47.4(6) | 57.5(3) | 56.7(6) | 51.8(4) | 46.9(4) | 61.9(7) | 63.4(5) | 49.3(4) |
| Ddet | 5.73 | 49 | 43.3(5) | 48.3(5) | 55.9(6) | 47.2(5) | 50.3(5) | 52.7(8) | 47.5(5) | 38.4(6) | 63.5(4) | 54.3(9) | 46.9(5) |
| PCdet | 6 | 48.7 | 40.6(6) | 42.1(9) | 61.7(5) | 55.4(4) | 43.6(8) | 58.5(4) | 44.8(6) | 46.3(5) | 52.0(9) | 64.9(4) | 42.6(6) |
| Struck | 7.18 | 44.4 | 35.4(7) | 47.0(7) | 53.4(9) | 45.0(7) | 43.9(9) | 58.0(5) | 39.0(7) | 30.4(10) | 63.5(3) | 54.4(8) | 40.6(7) |
| VTD | 7.64 | 43 | 30.9(10) | 48.8(4) | 53.9(8) | 38.6(9) | 46.2(6) | 57.3(7) | 37.2(8) | 28.3(11) | 63.1(6) | 54.9(7) | 38.5(8) |
| RGBdet | 9.36 | 39.9 | 26.7(13) | 40.9(10) | 54.7(7) | 31.9(12) | 46.0(7) | 50.5(10) | 35.7(9) | 34.8(7) | 46.8(11) | 56.2(6) | 33.7(11) |
| PCflow | 10.82 | 37.1 | 35.2(8) | 29.1(14) | 43.6(11) | 42.2(8) | 33.2(13) | 47.2(12) | 33.1(10) | 32.4(9) | 43.5(12) | 41.3(13) | 35.5(9) |
| CT | 10.91 | 36.4 | 31.1(11) | 46.7(8) | 36.9(13) | 39.0(10) | 34.4(12) | 48.6(11) | 31.5(11) | 23.3(14) | 54.3(8) | 42.1(12) | 34.2(10) |
| MIL | 11.18 | 35.5 | 29.0(12) | 35.1(12) | 44.4(10) | 32.5(13) | 38.5(10) | 51.6(9) | 29.7(13) | 33.8(8) | 38.7(13) | 50.2(10) | 30.5(13) |
| TLD | 11.55 | 35.9 | 32.2(9) | 37.2(11) | 38.3(12) | 36.6(11) | 34.6(11) | 45.5(13) | 31.5(12) | 25.6(12) | 49.0(10) | 40.4(14) | 33.6(12) |
| SemiB | 13.55 | 32.3 | 22.5(14) | 33.0(13) | 32.7(14) | 24.0(14) | 31.6(14) | 38.2(14) | 24.4(14) | 25.1(13) | 32.7(14) | 41.9(11) | 23.2(14) |
| OF | 15.27 | 18.6 | 17.9(15) | 11.4(16) | 23.4(15) | 20.1(15) | 17.5(16) | 18.1(16) | 18.8(15) | 15.9(15) | 22.3(15) | 23.4(15) | 16.8(15) |
| IIDfirstBB | 15.27 | 15.1 | 9.7(16) | 20.9(14) | 18.3(15) | 13.3(16) | 16.5(14) | 31.2(14) | 8.7(16) | 15.0(16) | 15.3(16) | 13.6(16) | 19.2(15) |
| IIDrandSize | 17.18 | 5.1 | 2.8(17) | 7.0(18) | 6.8(17) | 4.0(18) | 6.0(17) | 10.9(17) | 2.8(17) | 5.0(17) | 5.2(17) | 5.1(17) | 5.2(17) |
| IIDcenterBB | 17.91 | 4.6 | 0.7(19) | 9.4(17) | 6.5(18) | 4.6(17) | 4.6(18) | 10.0(18) | 2.4(18) | 4.8(18) | 4.3(18) | 5.1(18) | 3.3(18) |
| IIDrandLoc | 19 | 2 | 2.3(18) | 1.2(19) | 2.1(19) | 3.0(19) | 1.2(20) | 2.0(19) | 2.0(19) | 2.0(19) | 2.0(19) | 2.0(19) | 2.2(19) |
| IIDrandLocSize | 20 | 0.7 | 0.6(20) | 0.4(21) | 0.8(20) | 0.8(20) | 0.6(19) | 0.9(20) | 0.6(20) | 0.7(20) | 0.7(20) | 0.7(20) | 0.6(20) |
| IIDrandWithoutPrior | 20.91 | 0.1 | 0.1(21) | 0.0(20) | 0.1(21) | 0.1(21) | 0.0(21) | 0.1(21) | 0.1(21) | 0.1(21) | 0.1(21) | 0.1(21) | 0.1(21) |

### Our Baseline Algorithms

Table 1: Strong RGBD baseline algorithms

| | |
|---|---|
| OF | Uses only optical flow tracking. |
| Ddet | Uses Depth HOG detection tracking. |
| RGBdet | Uses RGB HOG detection tracking. |
| RGB +OF | Uses RGB HOG detection with optical flow. |
| RGBD +OF | Uses RGBD HOG detection with optical flow. |
| RGBDOcc +OF | Uses RGBD HOG detection and optical flow with occlusion handling. |
| PCflow | Uses 3D point tracker. |
| PCdet | Uses point cloud detection. |
| PC(det+flow) | Uses point cloud detection and 3D point tracking with occlusion handling. |

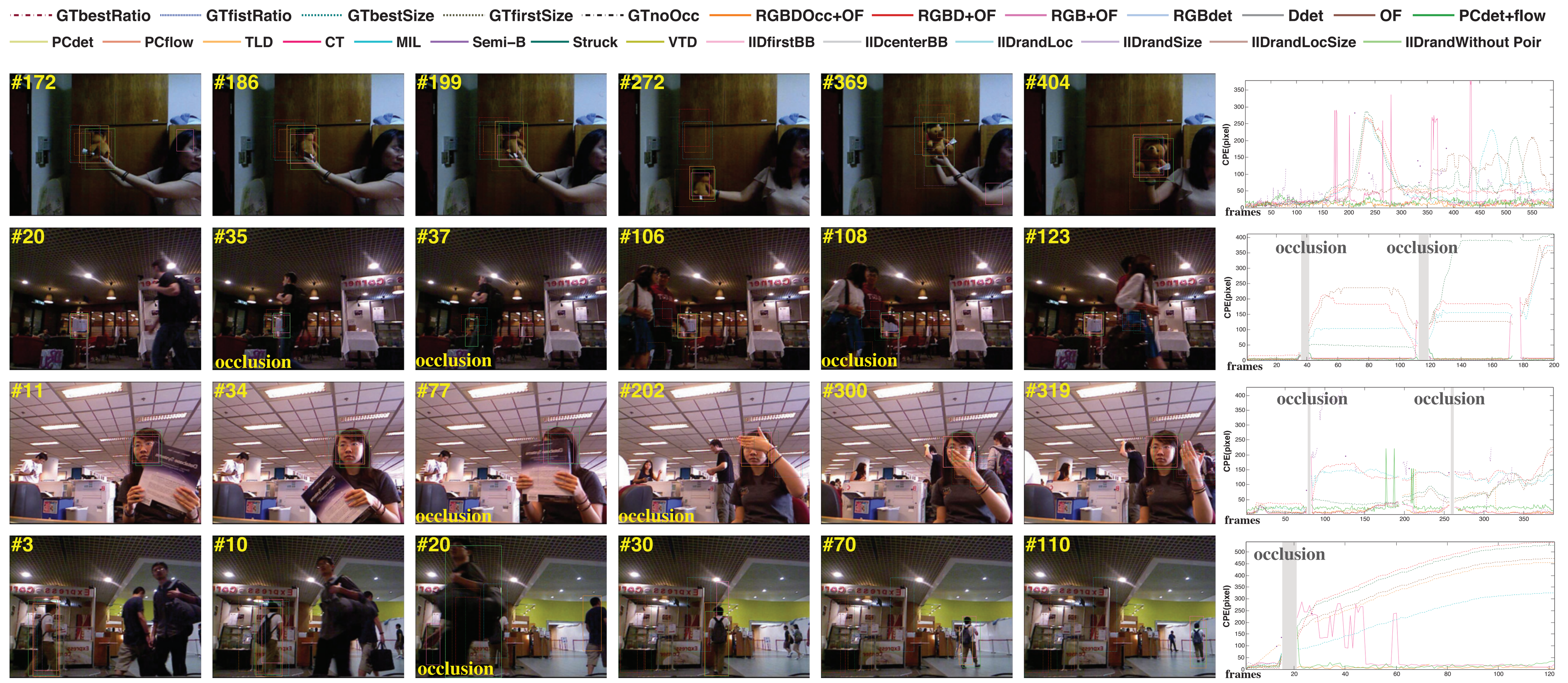
Table 2: Performance upper-bounds (GT:ground truth)

| | |
|---|---|
| GTfirstSize | Uses the GT location and first frame box size. |
| GTbestsize | Uses the GT location and fixed box size that optimize the successful rate. |
| GTfirstRatio | Uses the GT location and first frame box aspect ratio. |
| GTbestRatio | Uses the GT location and fixed box aspect ratio that optimize the successful rate. |
| GTnoOcc | Outputs the GT box, if exists, and a random box otherwise. |

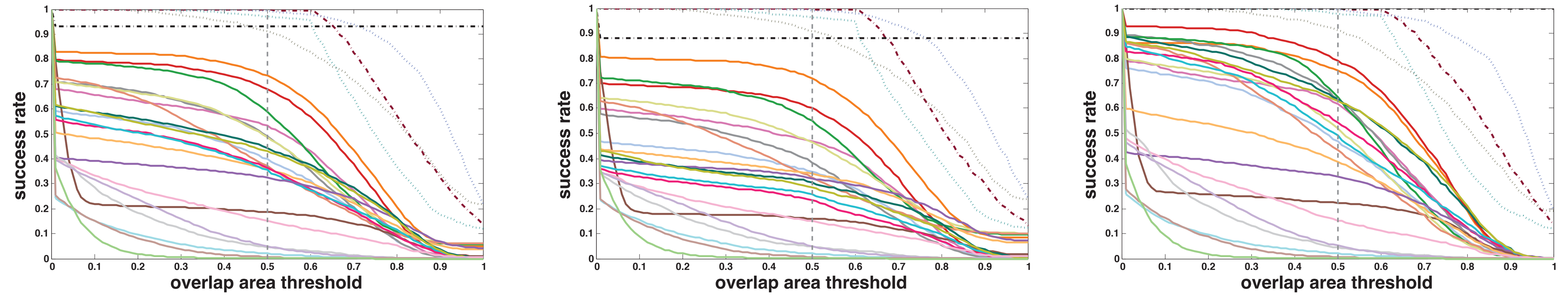
Table 3: Performance lower-bound algorithms

| | |
|---|---|
| IIDfirstBB | Always outputs the first frame bounding box for all frames. |
| IIDcenterBB | Always outputs the box locate at center of image, with first frame box size. |
| IIDrandSize | Outputs bounding boxes with the first frame box location and a random size based on dataset statistics. |
| IIDrandLoc | Outputs bounding boxes with the first frame box size and a random location based on dataset statistics |
| IIDrandLoc Size | Outputs bounding boxes with random location and size based on dataset statistics |
| IIDrand WithoutPrior | Outputs bounding boxes with random location and size without any prior knowledge of dataset |

## Results




## Success Rate




Average Error Rate