TLD dataset

Zdenek Kalal
zdenek.kalal@gmail.com
17.12.2010

The TLD dataset is a collection of 10 video sequences and an evaluation protocol for comparison of object tracking algorithms. Figure 1 shows snapshots from the sequences and table 1 lists their properties. The sequences were manually annotated with bounding boxes. More than 50% of occlusion or more than 90 degrees of out-of-plane rotation was annotated as "not visible".

If you use this dataset, please cite our CVPR'10 paper:

```
@article{Kalal2010,
 author = {Kalal, Z and Matas, J and Mikolajczyk, K},
 journal = {Conference on Computer Vision and Pattern Recognition},
 title = {{P-N Learning: Bootstrapping Binary Classifiers by Structural
Constraints}},
 year = {2010}
}
```

Name	Frames	Mov. camera	Partial occ.	Full occ.	Pose change	Illum. change	Scale change	Similar objects
1. David	761	yes	yes	no	yes	yes	yes	no
2. Jumping	313	yes	no	no	no	no	no	no
3. Pedestrian 1	140	yes	no	no	no	no	no	no
4. Pedestrian 2	338	yes	yes	yes	no	no	no	yes
5. Pedestrian 3	184	yes	yes	yes	no	no	no	yes
6. Car	945	yes	yes	yes	no	no	no	yes
7. Motocross	2665	yes	yes	yes	yes	yes	yes	yes
8. Volkswagen	8576	yes	yes	yes	yes	yes	yes	yes
9. Carchase	9928	yes	yes	yes	yes	yes	yes	yes
10. Panda	3000	yes	yes	yes	yes	yes	yes	no

Table 1: Properties of sequences

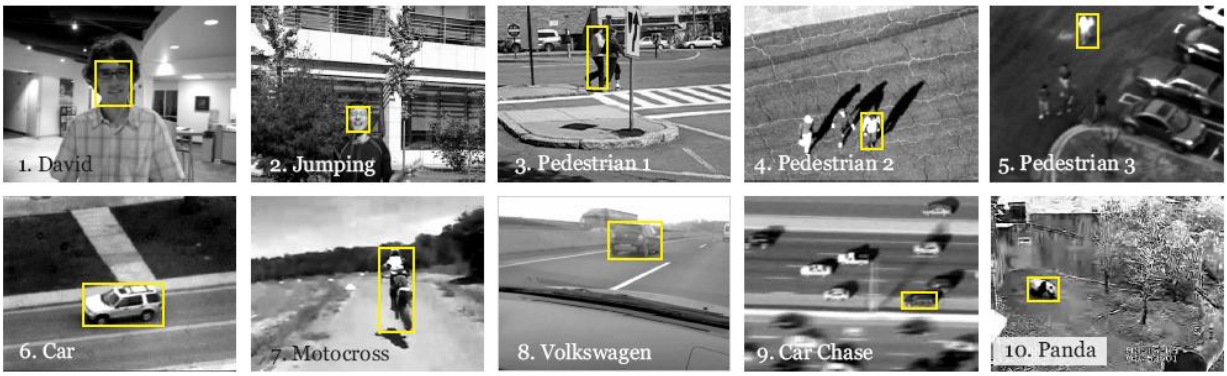

Fig.1: Snapshots from the sequences with the objects marked by the bounding box

Contents of the dataset

The dataset contains an evaluation script in directory "_matlab" and 10 data-directories containing:

- a sequence of images
- **gt.txt**, the ground truth trajectory,
- **init.txt**, the initial object's location,
- outputs of trackers:
  OnlineBoost [1], SemiBoost [2], BeyondSemiBoost [3], MIL [4], CoGD [5], TLD1.0 [6]

Object *state* has two possible values:

- bounding box: [left_column, top_row, right_column, bottom_row]
- not-visible : [nan, nan, nan, nan]

Evaluation protocol

The performance of tracking algorithms is evaluated using precision P, recall R and f-measure F. P is the number of true positives divided by number of all responses, R is the number true positives divided by the number of object occurrences that should have been detected. F combines these two measures as $F = 2PR/(P + R)$. Detection is considered to be correct if its overlap with ground truth bounding box was larger than 25% (see figure 2).

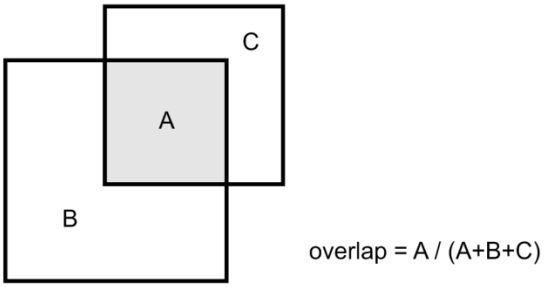

Fig. 2: Bounding box overlap

Normalization

When comparing a trajectory to ground truth, the trajectory is normalized (shift, aspect and scale correction) so that the first bounding box matches the ground truth; all remaining bounding boxes are normalized with the same parameters. Therefore, you are free to use it arbitrary initialization of your tracker.

References

1. Grabner H, Bischof H. On-line Boosting and Vision. *Conference on Computer Vision and Pattern Recognition*. 2006
2. Grabner H, Leistner C, Bischof H. Semi-Supervised On-line Boosting for Robust Tracking. *European Conference on Computer Vision*. 2008.
3. Stalder S, Grabner H, Gool LV. Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition. *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. 2009
4. Babenko B, Yang M-H, Belongie S. Visual Tracking with Online Multiple Instance Learning. *Conference on Computer Vision and Pattern Recognition*. 2009.
5. Yu Q, Dinh TB, Medioni G. Online tracking and reacquisition using co-trained generative and discriminative trackers. *European Conference on Computer Vision*. 2008.
6. Kalal Z, Matas J, Mikolajczyk K. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. *Conference on Computer Vision and Pattern Recognition*. 2010.