Struck: Structured Output Tracking with Kernels

Sam Hare, Amir Saffari, And Philip H. S. Torr

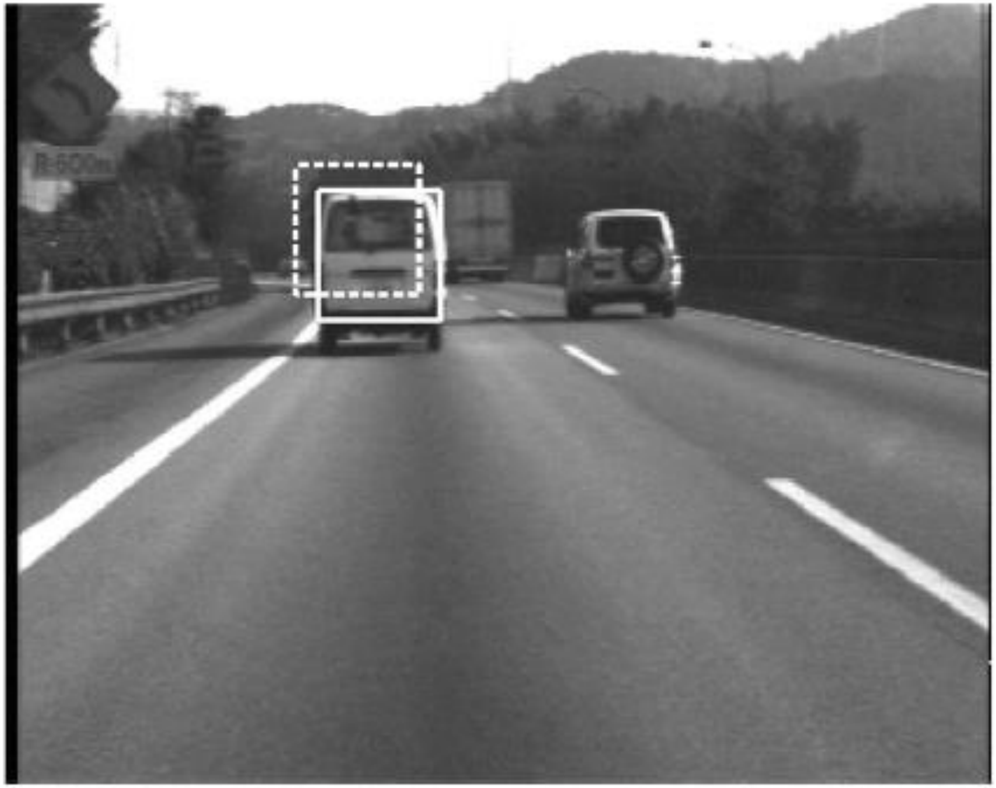
International Conference On Computer Vision (ICCV), 2011

Motivations

Problem: tracking-by-detection

Input: target

Output: locations over times



Performance summary

Struck

TLD

MIL

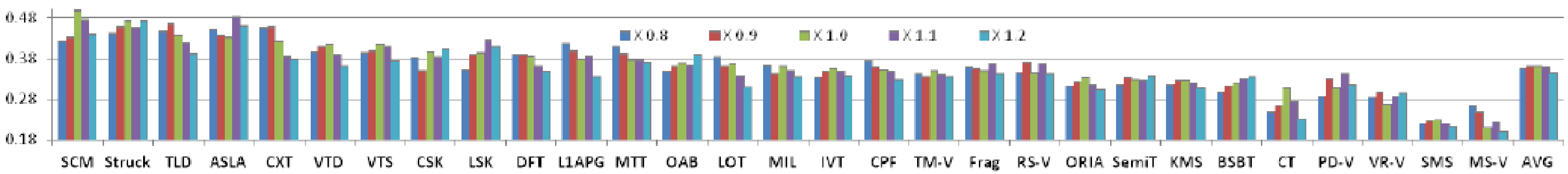


Figure 6. Performance summary for the trackers initialized with different size of bounding box. AVG (the last one) illustrates the average performance over all trackers for each scale.

Y Wu, J Lim, MH Yang "Online Object Tracking: A Benchmark", Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on

Outline

Previous works

- Tracking-by-detection
- Adaptive tracking-by-detection

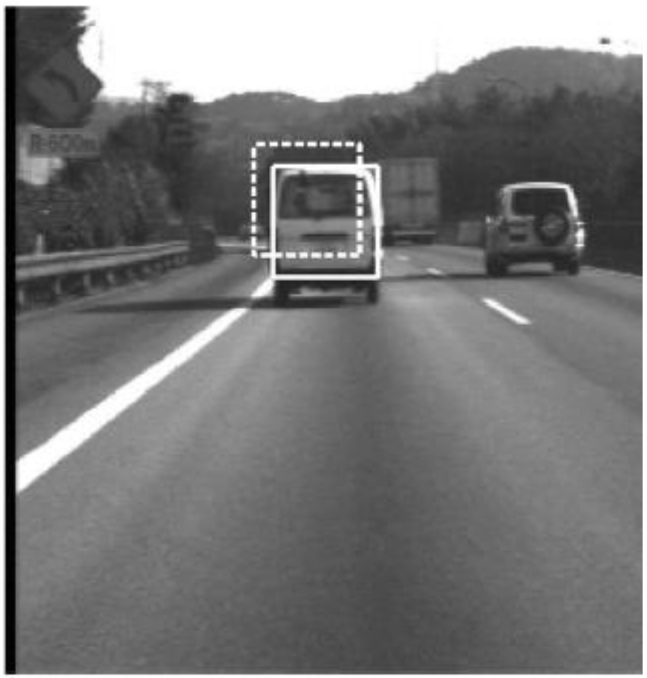
Methods

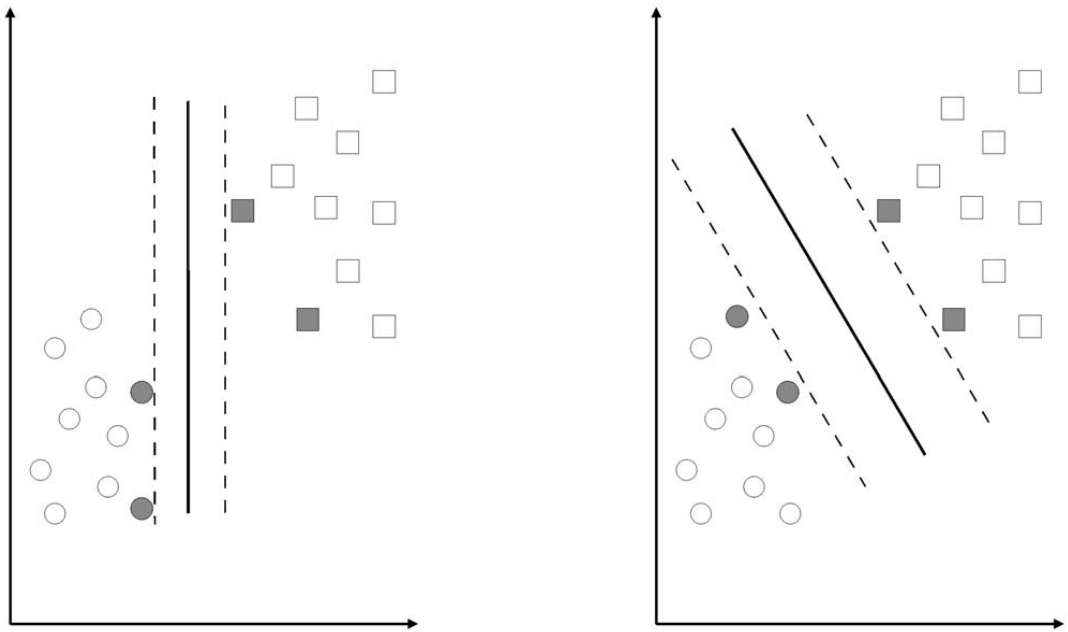
- Structured output tracking
- Online optimization and budget mechanism

Experiments and results

Previous Works

Tracking problem as a detection task applied over time





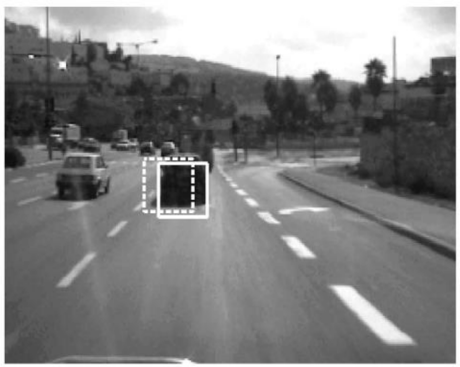
(a) (b)

Separating hyperplanes with different margins.

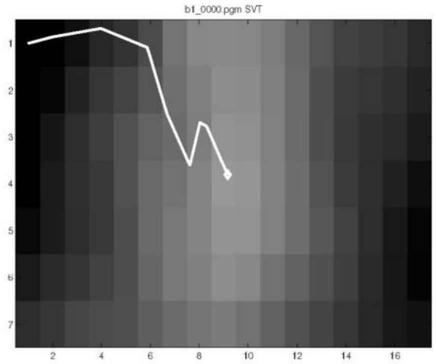
S. Avidan. Support Vector Tracking. IEEE Trans. on PAMI, 26:1064–1072, 2004.

Previous Works

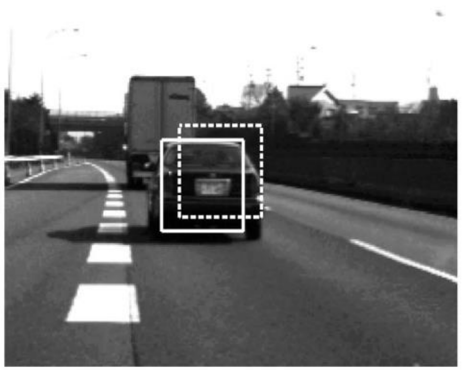
Tracking problem as a detection task applied over time



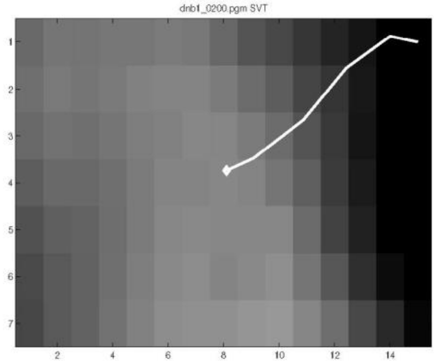
(a)



(b)



(c)



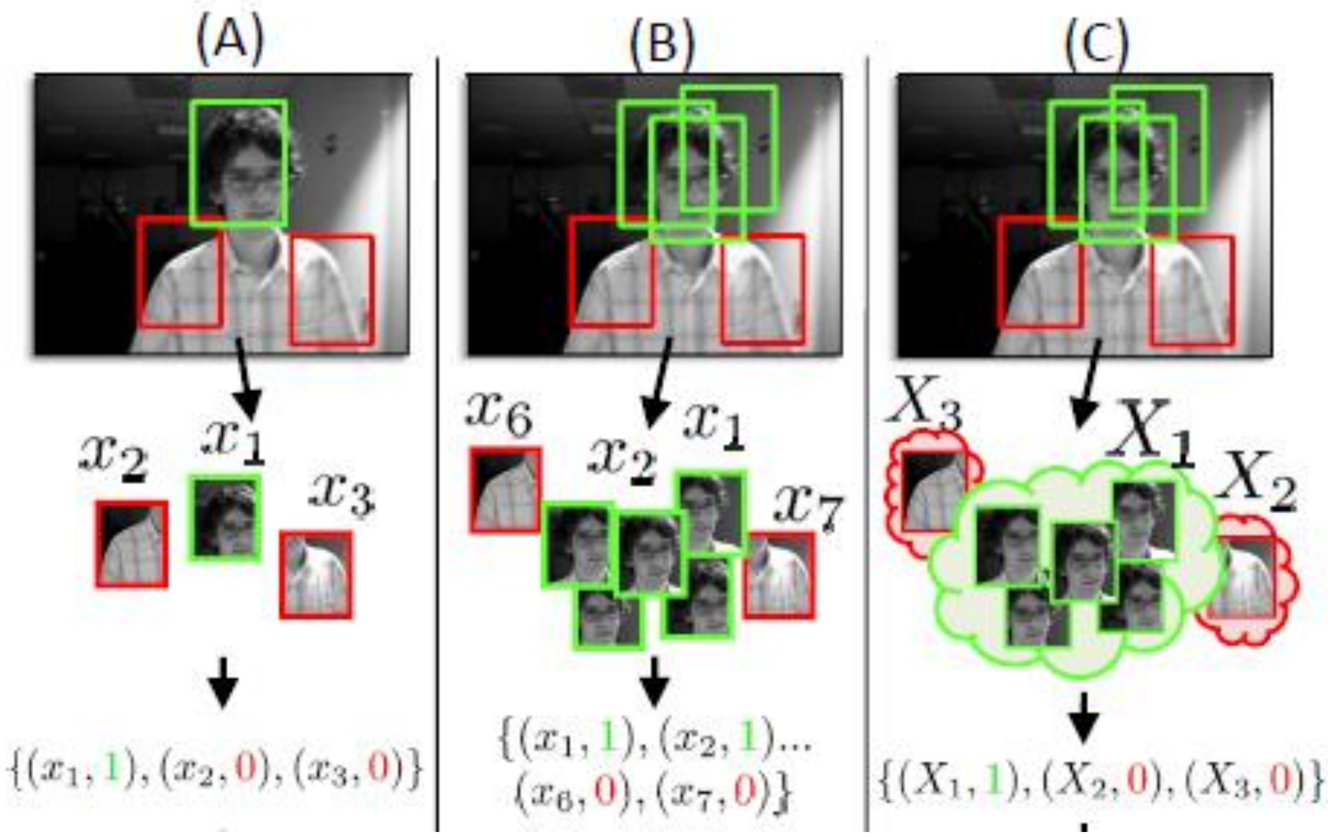
(d)

look for the image region with the highest SVM score

S. Avidan. Support Vector Tracking. IEEE Trans. on PAMI, 26:1064–1072, 2004.

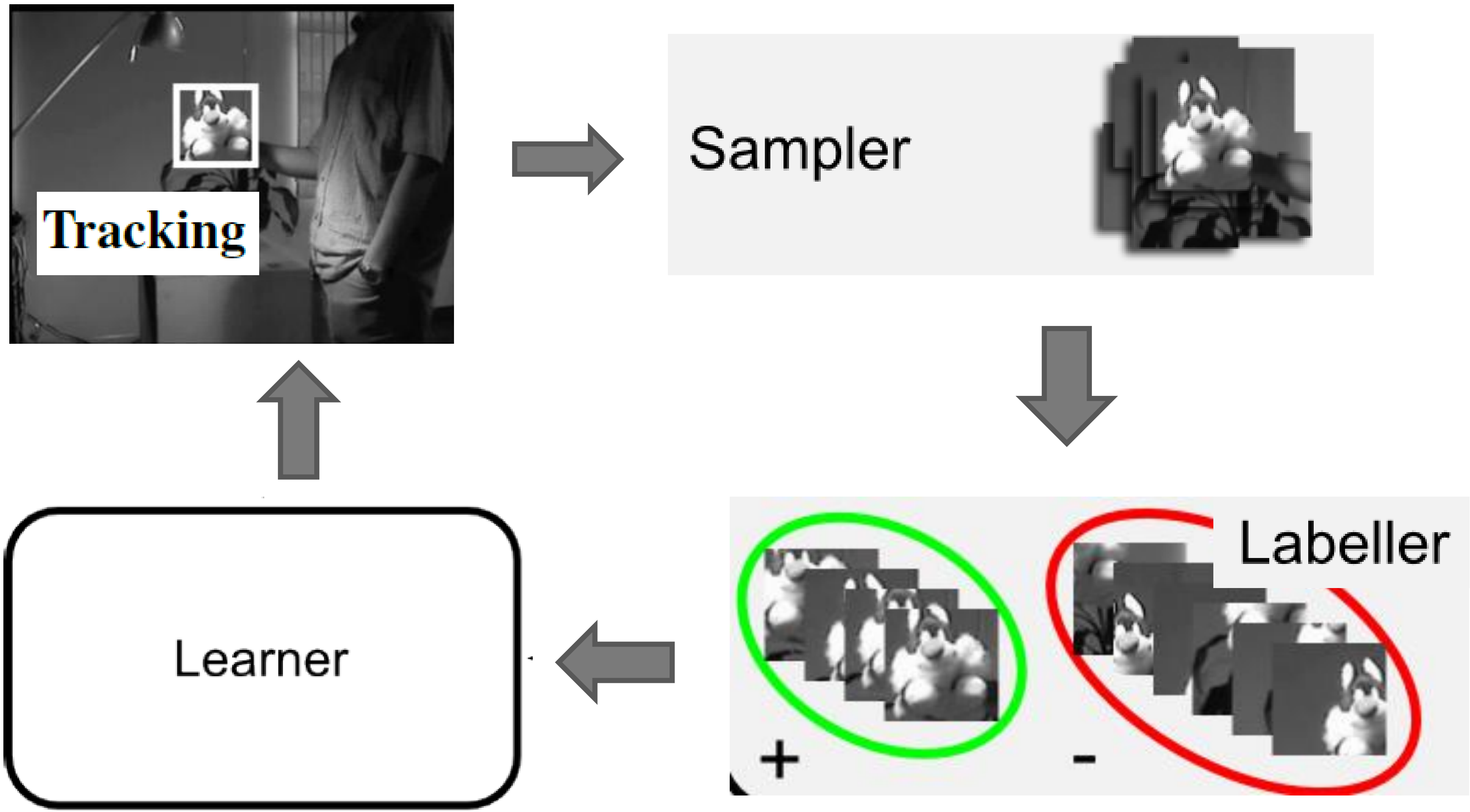
Previous Works

Adaptive tracking-by-detection



B. Babenko, M. H. Yang, and S. Belongie. Visual Tracking with Online Multiple Instance Learning. In Proc. CVPR, 2009.

Previous Works – Adaptive Tracking-by-detection

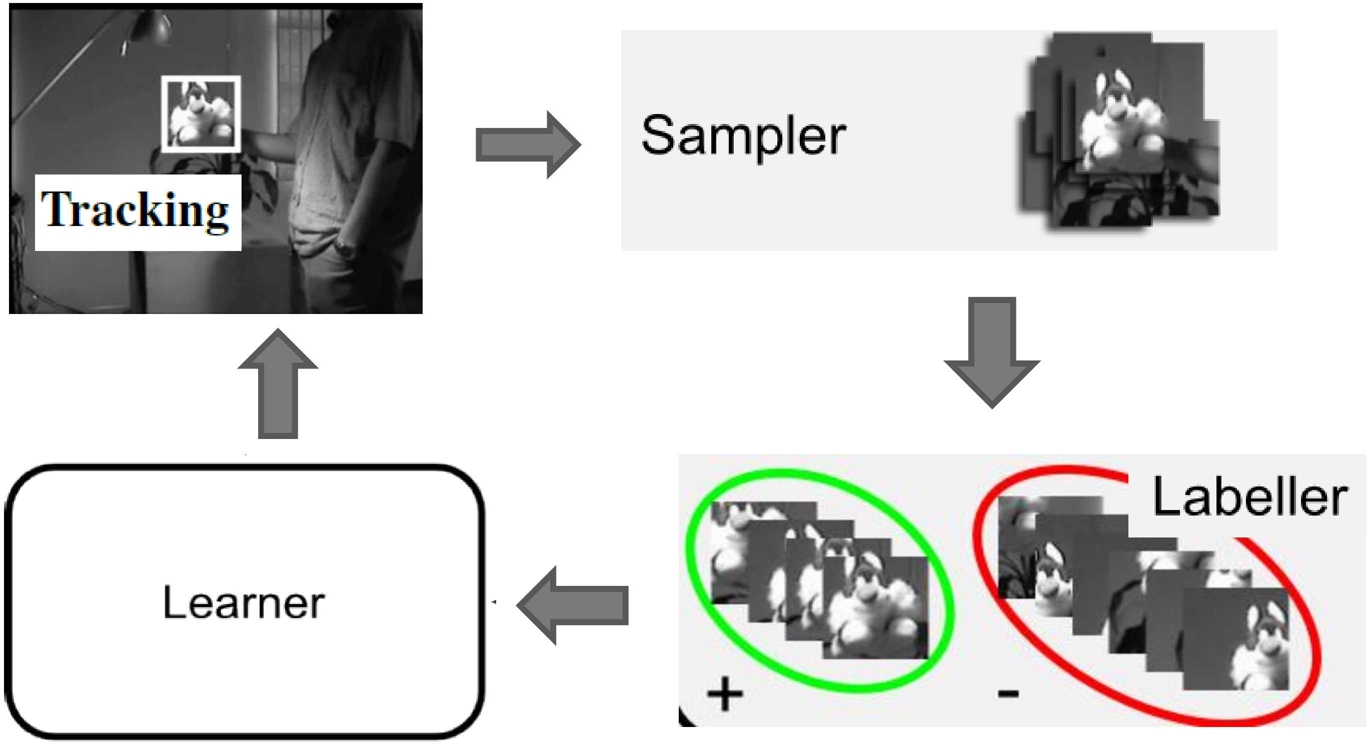


Previous Works – Adaptive Tracking-by-detection

Adaptive tracking-by-detection

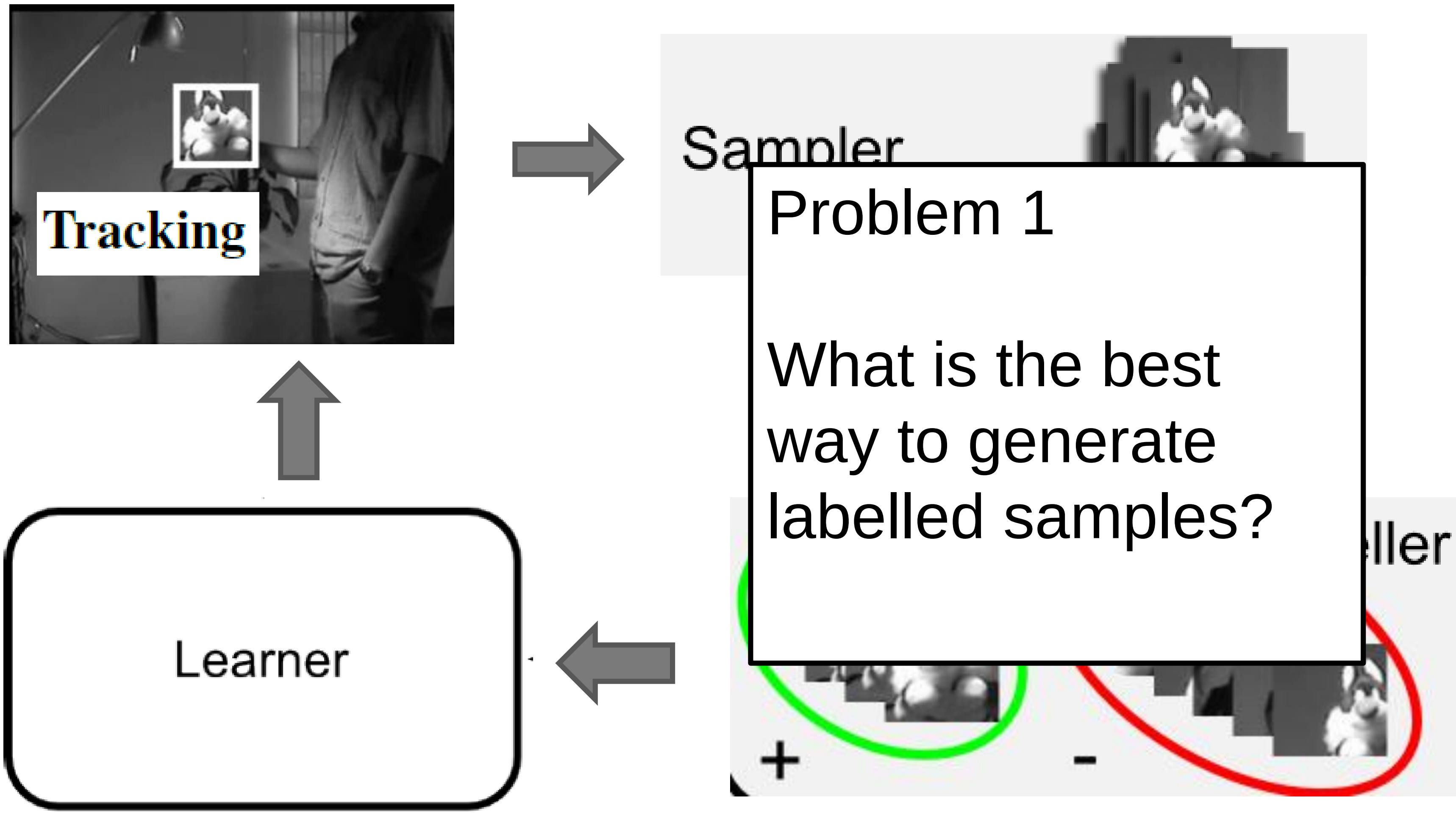
Tracking: A classification task

Learning: A update the object model.



B. Babenko, M. H. Yang, and S. Belongie. Visual Tracking with Online Multiple Instance Learning. In Proc. CVPR, 2009.

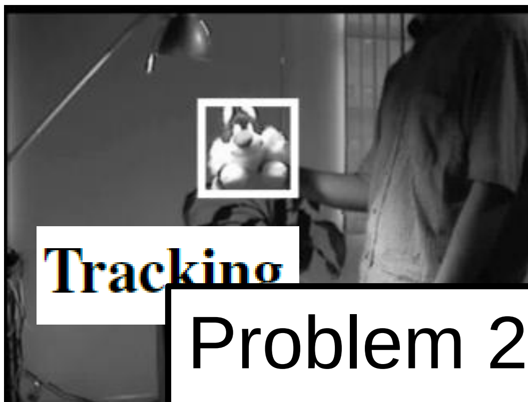
Previous Works – Adaptive Tracking-by-detection

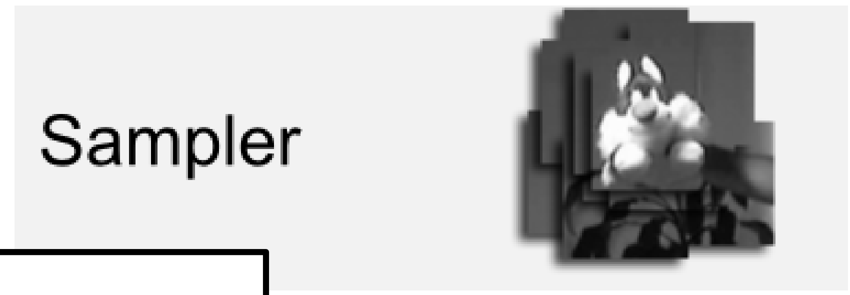


Problem 1

What is the best way to generate labelled samples?

Previous Works – Adaptive Tracking-by-detection

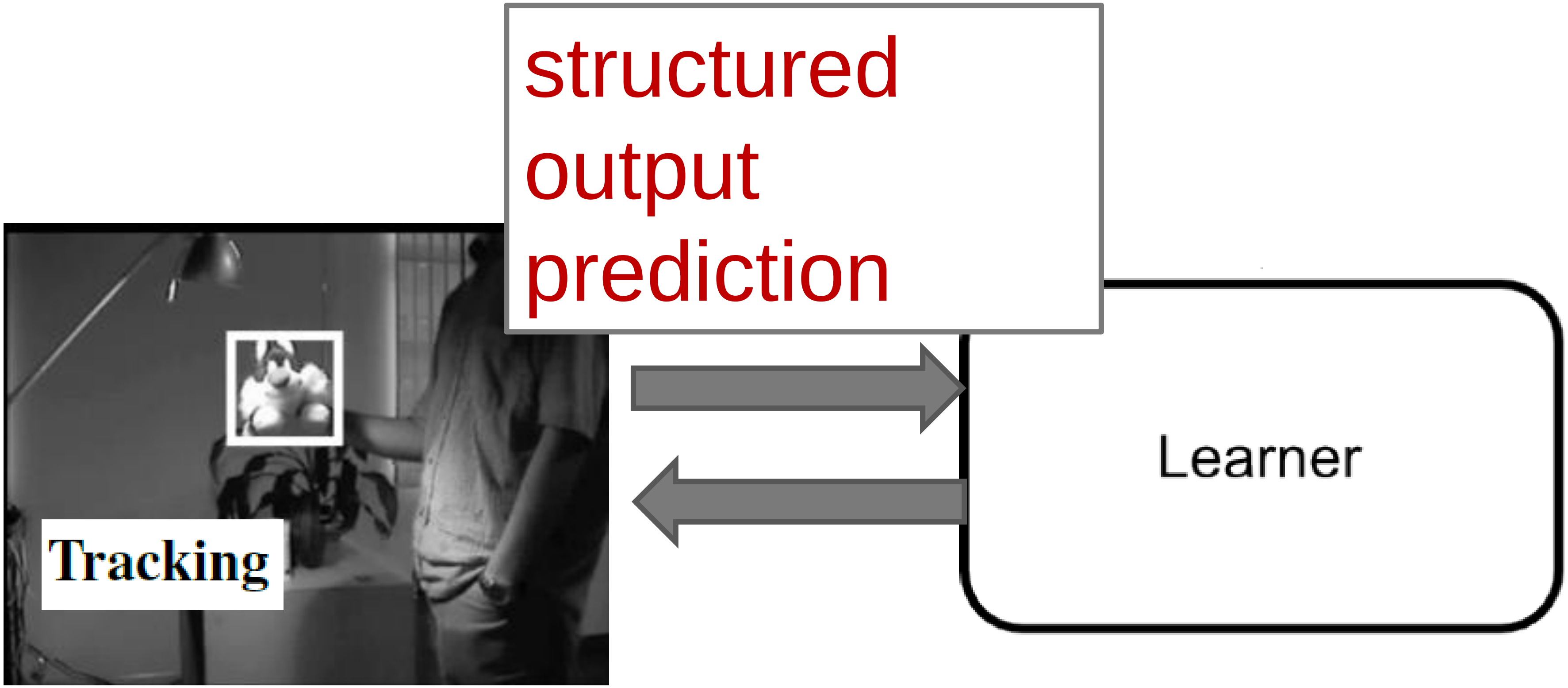




Problem 2

Label prediction and position estimation are different objectives.

Main Idea



Main Contributions

Structured output tracking

Avoid the intermediate classification step

Online learning and budgeting mechanism

Prevents too many training data

Outline

Previous work

- Tracking-by-detection
- Adaptive tracking-by-detection

Methods

- **Structured output tracking**
- Online optimization and budget mechanism

Experiments and results

Structured Output Tracking

tracker position

$$\mathbf{y}_t = f(\mathbf{x}_t^{\mathbf{p}_{t-1}}) = \arg\max_{\mathbf{y}\in\mathcal{Y}} F(\mathbf{x}_t^{\mathbf{p}_{t-1}}, \mathbf{y})$$

Best motions

search window

image patch

M. B. Blaschko and C. H. Lampert. Learning to Localize Objects with Structured Output Regression. In Proc. ECCV, 2008.

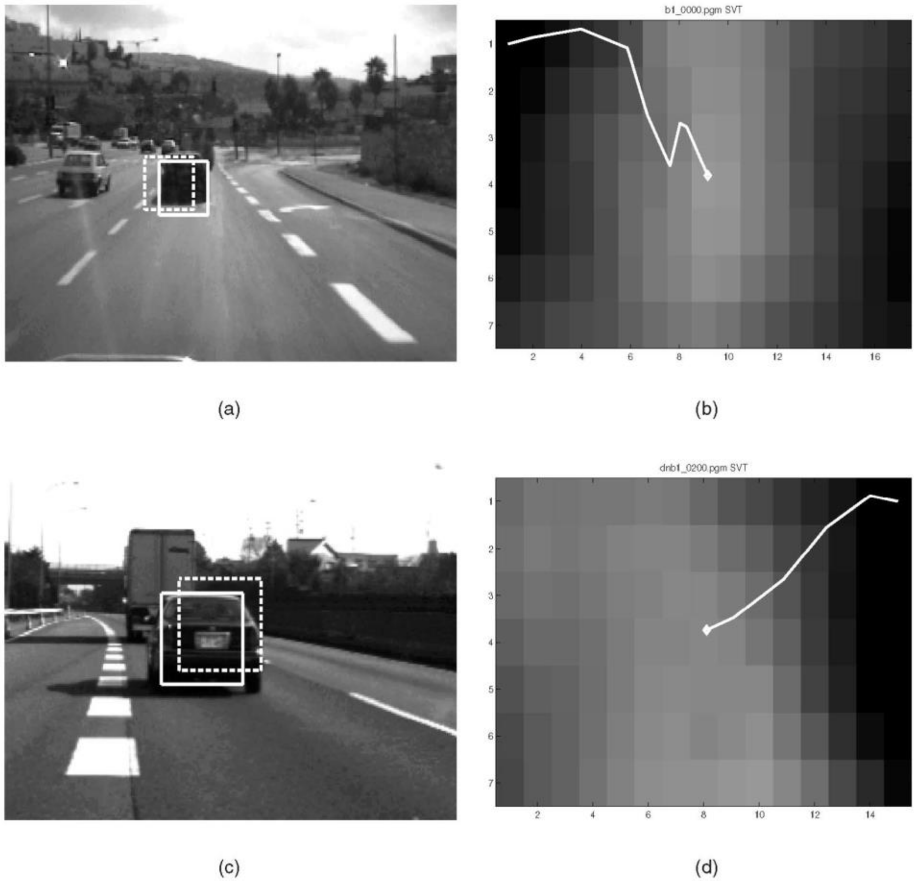
Structured Output Tracking

The output space is all transformations instead of the binary labels.

$$\mathbf{y}_t = f(\mathbf{x}_t^{\mathbf{p}_{t-1}}) = \arg\max_{\mathbf{y} \in \mathcal{Y}} F(\mathbf{x}_t^{\mathbf{p}_{t-1}}, \mathbf{y})$$

M. B. Blaschko and C. H. Lampert. Learning to Localize Objects with Structured Output Regression. In Proc. ECCV, 2008.

Structured SVM Model



The SVM score should correlate with overlapping size with the best tracking bounding box.

S. Avidan. Support Vector Tracking. IEEE Trans. on PAMI, 26:1064–1072, 2004.

Structured Output Tracking

Algorithm 2 Struck: Structured Output Tracking

Require: $\mathrm{f}_t, \mathrm{p}_{t-1}, \mathcal{S}_{t-1}$

Estimate change in object location

$\mathbf{y}_t = \arg\max_{\mathbf{y}\in\mathcal{Y}} F(\mathrm{x}_t^{\mathrm{p}_{t-1}}, \mathbf{y})$

$\mathrm{p}_t = \mathrm{p}_{t-1} \circ \mathbf{y}_t$

Update discriminant function

$(i, \mathbf{y}_+, \mathbf{y}_-) \leftarrow \textsc{ProcessNew}(\mathrm{x}_t^{\mathrm{p}_t}, \mathbf{y}^0)$

$\textsc{SMOStep}(i, \mathbf{y}_+, \mathbf{y}_-)$

$\textsc{BudgetMaintenance}()$

for $j = 1$ to n_R **do**

  $(i, \mathbf{y}_+, \mathbf{y}_-) \leftarrow \textsc{ProcessOld}()$

  $\textsc{SMOStep}(i, \mathbf{y}_+, \mathbf{y}_-)$

  $\textsc{BudgetMaintenance}()$

  for $k = 1$ to n_O **do**

    $(i, \mathbf{y}_+, \mathbf{y}_-) \leftarrow \textsc{Optimize}()$

    $\textsc{SMOStep}(i, \mathbf{y}_+, \mathbf{y}_-)$

  end for

end for

return $\mathrm{p}_t, \mathcal{S}_t$

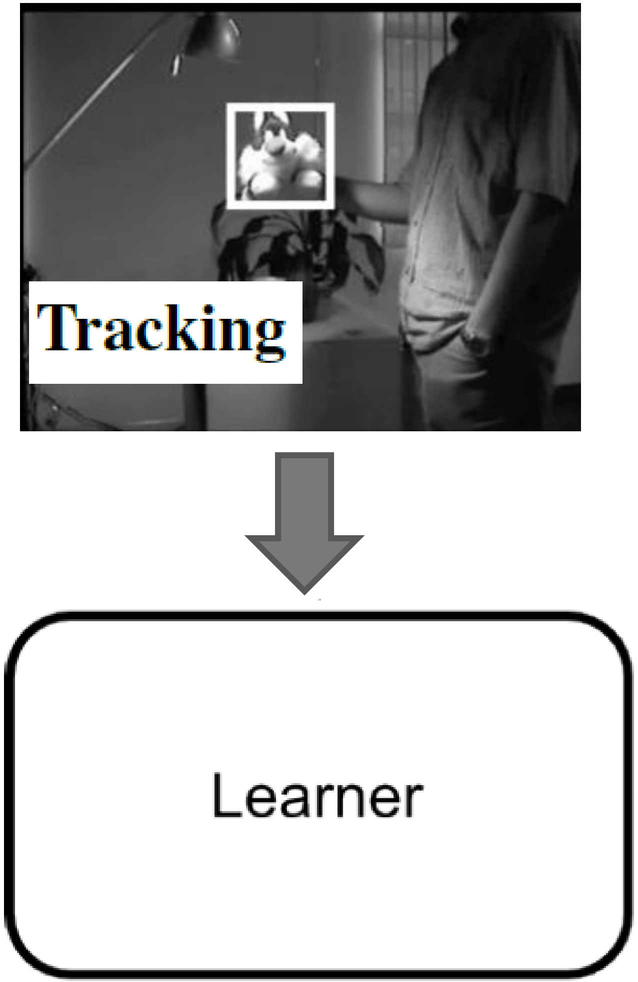
Structured Output Tracking

Algorithm 2 Struck: Structured Output Tracking

Require: $\mathrm{f}_t, \mathrm{p}_{t-1}, \mathcal{S}_{t-1}$

Estimate change in object location
$\mathbf{y}_t = \arg\max_{\mathbf{y} \in \mathcal{Y}} F(\mathbf{x}_t^{\mathrm{p}_{t-1}}, \mathbf{y})$
$\mathrm{p}_t = \mathrm{p}_{t-1} \circ \mathbf{y}_t$
Update discriminant function
$(i, \mathbf{y}_+, \mathbf{y}_-) \leftarrow$ PROCESSNEW$(\mathbf{x}_t^{\mathrm{p}_t}, \mathbf{y}^0)$
SMOSTEP$(i, \mathbf{y}_+, \mathbf{y}_-)$
BUDGETMAINTENANCE()
for $j = 1$ to n_R **do**
    $(i, \mathbf{y}_+, \mathbf{y}_-) \leftarrow$ PROCESSOLD()
    SMOSTEP$(i, \mathbf{y}_+, \mathbf{y}_-)$
    BUDGETMAINTENANCE()
    for $k = 1$ to n_O **do**
        $(i, \mathbf{y}_+, \mathbf{y}_-) \leftarrow$ OPTIMIZE()
        SMOSTEP$(i, \mathbf{y}_+, \mathbf{y}_-)$
    end for
end for
return $\mathrm{p}_t, \mathcal{S}_t$

Structured Output Tracking



Algorithm 2 Struck: Structured Output Tracking

Require: $\mathrm{f}_t, \mathrm{p}_{t-1}, \mathcal{S}_{t-1}$

Estimate change in object location

$\mathrm{y}_t = \arg\max_{\mathrm{y}\in\mathcal{Y}} F(\mathrm{x}_t^{\mathrm{p}_{t-1}}, \mathrm{y})$

$\mathrm{p}_t = \mathrm{p}_{t-1} \circ \mathrm{y}_t$

Update discriminant function

$(i, \mathrm{y}_+, \mathrm{y}_-) \leftarrow$ PROCESSNEW$(\mathrm{x}_t^{\mathrm{p}_t}, \mathrm{y}^0)$

SMOSTEP$(i, \mathrm{y}_+, \mathrm{y}_-)$

BUDGETMAINTENANCE()

fo

Come back later

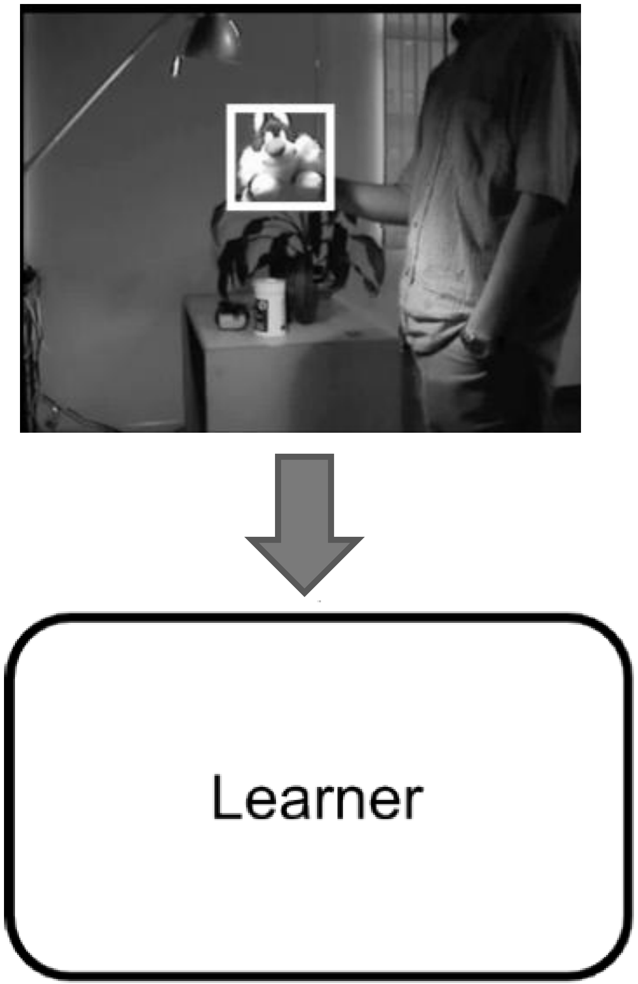
SMOSTEP$(i, \mathrm{y}_+, \mathrm{y}_-)$

end for

end for

return $\mathrm{p}_t, \mathcal{S}_t$

Structured output SVM



$$\mathbf{y}_t = f(\mathbf{x}_t^{\mathbf{p}_{t-1}}) = \arg\max_{\mathbf{y} \in \mathcal{Y}} F(\mathbf{x}_t^{\mathbf{p}_{t-1}}, \mathbf{y})$$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & \forall i: \ \xi_i \geq 0 \\ & \forall i, \forall \mathbf{y} \neq \mathbf{y}_i: \ \langle \mathbf{w}, \delta\Phi_i(\mathbf{y}) \rangle \geq \Delta(\mathbf{y}_i, \mathbf{y}) - \xi_i \end{aligned}$$

Efficient SMO optimization (CS229, EE364)
Kernels (CS229)

A. Bordes, L. Bottou, P. Gallinari, and J. Weston. Solving multiclass support vector machines with LaRank. In Proc. ICML, 2007.

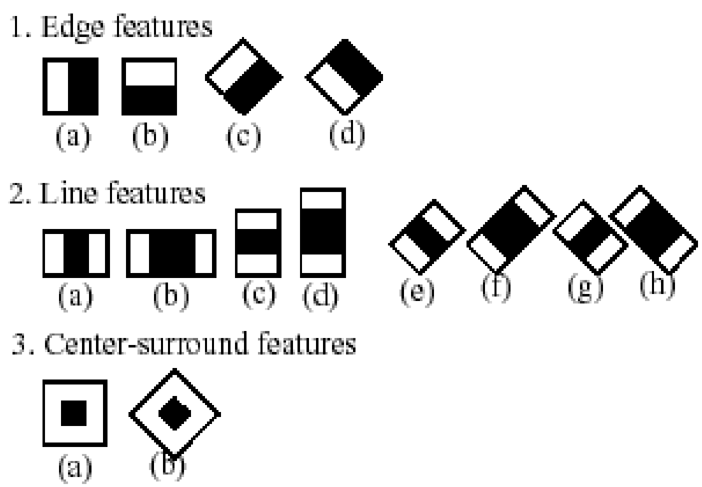
Structured output SVM

Gaussian kernel between image feature vectors (CS229)

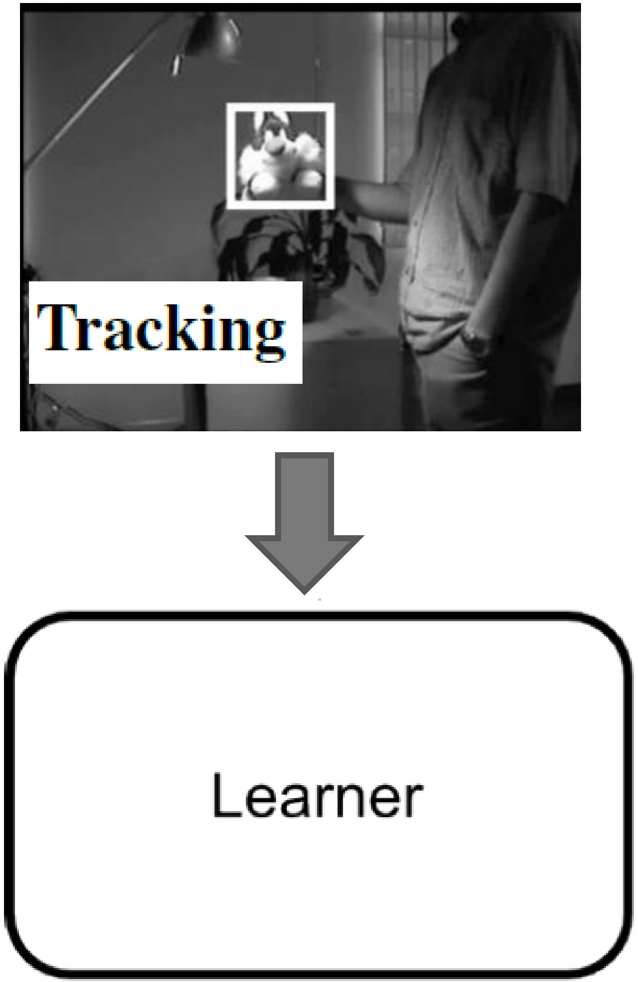
$$k(\mathrm{x}, \bar{\mathrm{x}}) = \exp(-\sigma \|\mathrm{x} - \bar{\mathrm{x}}\|^2),$$

Haar-like features (CS231A, CS232)

The responses of the Haar features are the input vectors of the kernel



Online optimization



Algorithm 2 Struck: Structured Output Tracking

Require: $\mathrm{f}_t, \mathrm{p}_{t-1}, \mathcal{S}_{t-1}$

Estimate change in object location

$\mathrm{y}_t = \arg\max_{\mathrm{y} \in \mathcal{Y}} F(\mathrm{x}_t^{\mathrm{p}_{t-1}}, \mathrm{y})$

$\mathrm{p}_t = \mathrm{p}_{t-1} \circ \mathrm{y}_t$

Update discriminant function

$(i, \mathrm{y}_+, \mathrm{y}_-) \leftarrow$ ProcessNew$(\mathrm{x}_t^{\mathrm{p}_t}, \mathrm{y}^0)$

SMOStep$(i, \mathrm{y}_+, \mathrm{y}_-)$

BudgetMaintenance()

for $j = 1$ to n_R **do**

  $(i, \mathrm{y}_+, \mathrm{y}_-) \leftarrow$ ProcessOld()

  SMOStep$(i, \mathrm{y}_+, \mathrm{y}_-)$

  BudgetMaintenance()

  for $k = 1$ to n_O **do**

    $(i, \mathrm{y}_+, \mathrm{y}_-) \leftarrow$ Optimize()

    SMOStep$(i, \mathrm{y}_+, \mathrm{y}_-)$

  end for

end for

return $\mathrm{p}_t, \mathcal{S}_t$

Outline

Previous work

- Tracking-by-detection
- Adaptive tracking-by-detection

Methods

- Structured output tracking
- **Online optimization and budget mechanism**

Experiments and results

Online optimization

PROCESSNEW():

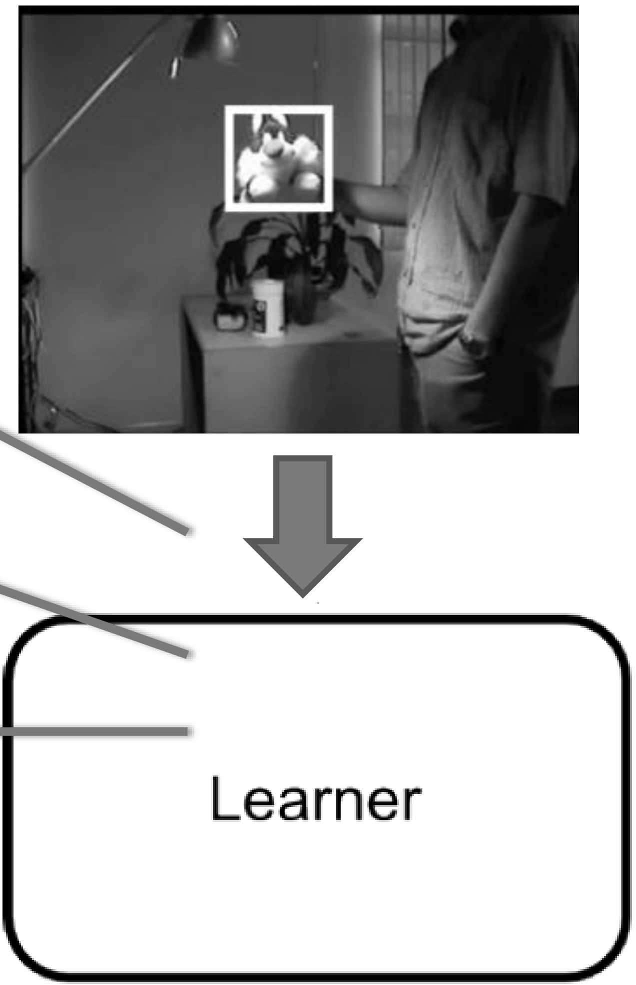
- Processes a new example

PROCESSOLD():

- Processes an existing support pattern

OPTIMIZE():

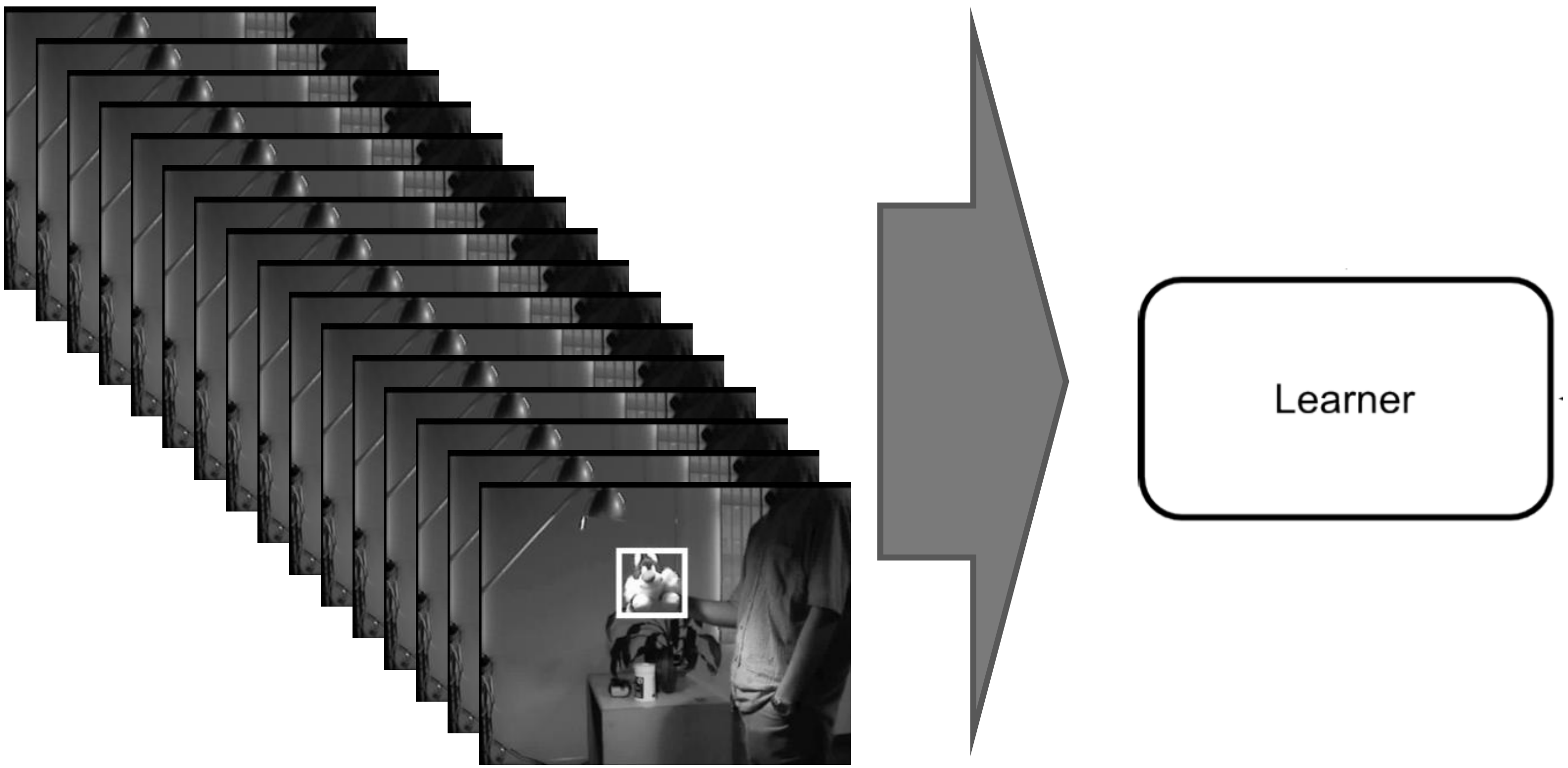
- Processes an existing support pattern chosen at random



Budget mechanism

The number of support vectors increase over time.

Computational and storage costs grow linearly with the number of support vectors.



Incorporating a budget

A budget (limit) of the number of supporting vectors.

Remove the support vector which results in the smallest change to the weight vector

K. Crammer, J. Kandola, R. Holloway, and Y. Singer. Online Classification on a Budget. In NIPS, 2003.
Z. Wang, K. Crammer, and S. Vucetic. Multi-Class Pegasos on a Budget. In Proc. ICML, 2010. 2

Outline

Previous works

- Tracking-by-detection
- Adaptive tracking-by-detection

Methods

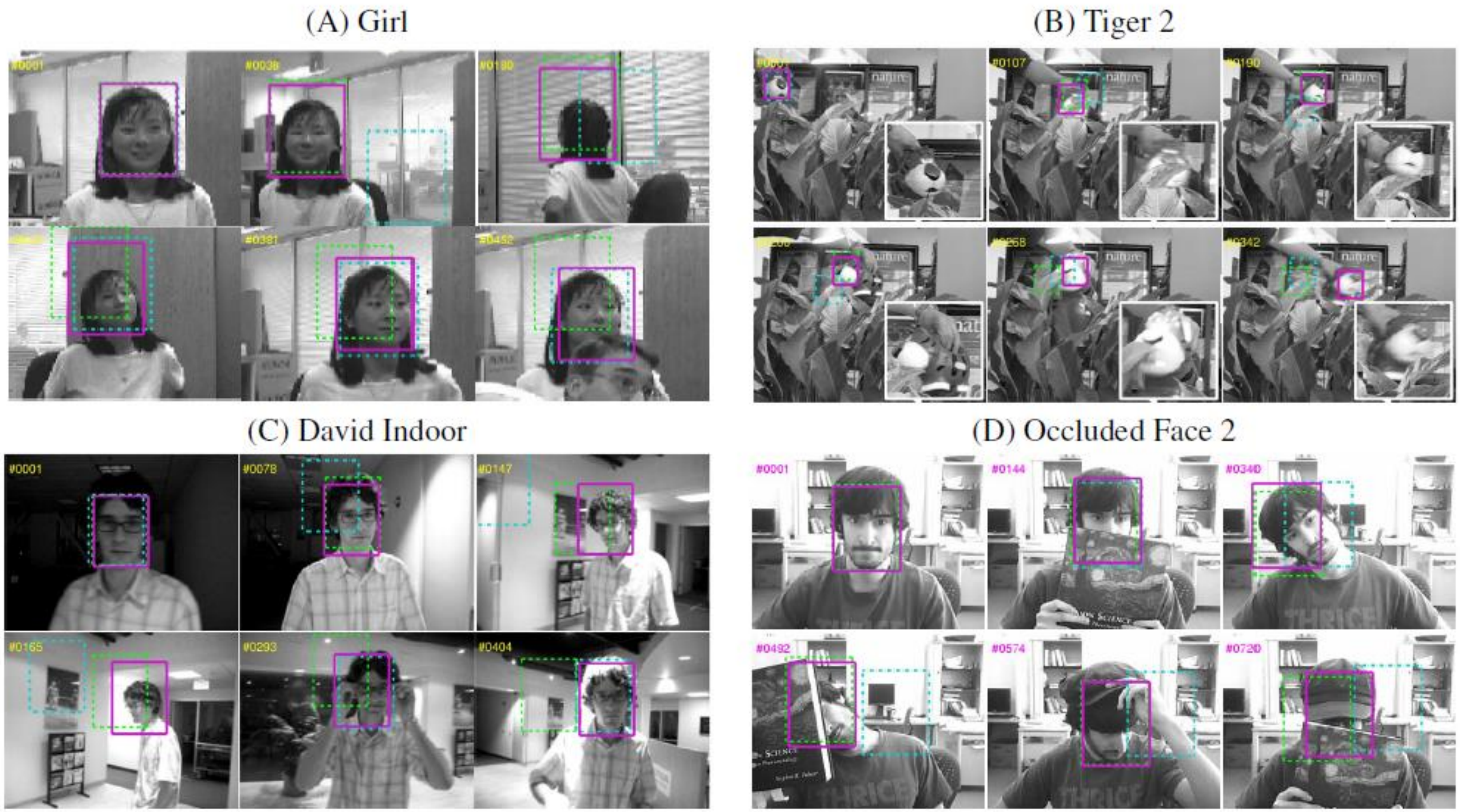
- Structured output tracking
- Online optimization and budget mechanism

Experiments and results

Experiments

- Haar-like features
  - 6 different types arranged on a grid at 2 scales on a 4 x 4 grid, resulting in 192 features
- Search radius 60, 5 radial and 16 angular divisions.
- Budget size is as low as B = 20, 50, 100, inf.

Dataset



http://vision.ucsd.edu/˜bbabenko/project_miltrack.shtml;
B. Babenko, M. H. Yang, and S. Belongie. Visual Tracking with Online Multiple Instance Learning. In Proc. CVPR, 2009.

Overlap criterion

Jaccard similarity of bounding boxes

$$s^{o}_{\mathbf{p}_t}(\mathbf{y}^i_t, \mathbf{y}^j_t) = \frac{(\mathbf{p}_t \circ \mathbf{y}^i_t) \cap (\mathbf{p}_t \circ \mathbf{y}^j_t)}{(\mathbf{p}_t \circ \mathbf{y}^i_t) \cup (\mathbf{p}_t \circ \mathbf{y}^j_t)}$$

Metric 2: Jaccard Similarity

$$100 \cdot \frac{\sum_{i=1}^{n}[\alpha_i = 1 \wedge f_i = 1]}{\sum_{i=1}^{n}[\alpha_i = 1 \vee f_i = 1]}$$

α_i: estimated foreground/background label
f_i: ground truth foreground/background label

Results

http://www.samhare.net/research/struck

Visualization of the support vector set

(a) *girl* (b) *david*

Comparison

http://www.samhare.net/research/struck

Results

Struck with the smallest budget size (B = 20) outperforms the state-of-the-art.

Average frames per second: 12 – 21.

Extensions

- Used more objection representations
  - Haar-like features
  - Raw pixel features
  - Histogram features
- Combining multiple kernels seems to improve results, but not significantly.

- Use key points and associated descriptors for object detection.
- Consider other machine learning algorithms.

Main Contributions

Structured output tracking

Avoid the intermediate classification step

Online learning and budgeting mechanism

Prevents too many training data

References

Sam Hare, Amir Saffari Philip H. S. Torr Struck: Structured Output Tracking with Kernels International Conference on Computer Vision (ICCV), 2011

A. Bordes, L. Bottou, P. Gallinari, and J. Weston. Solving multiclass support vector machines with LaRank. In Proc. ICML, 2007.

I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large Margin Methods for Structured and Interdependent Output Variables. JMLR, 6:1453–1484, Dec. 2005.

K. Crammer, J. Kandola, R. Holloway, and Y. Singer. Online Classification on a Budget. In NIPS, 2003.

P. Viola and M. J. Jones. Robust real-time face detection. IJCV, 57:137–154, 2004.

Thank you!