上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Visual Tracking via Structural Patch-based Dictionary Pair Learning

Tao Zhou[a], Fanghui Liu[a], Harish Bhaskar[b] , Jie Yang[a,*], Lei Chen[c], Ping Cai[d]

[a] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China.
[b] Dept. of Elec. & Comp. Engg., Khalifa Univ. of Science Technology and Research, U.A.E
[c] School of Computer Science, Nanjing University of Posts and Telecommunications, China.
[d] Department of Instrument Science and Engineering, Shanghai Jiao Tong University, China.

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Outline

- Introduction & Motivation
- Proposed method
- Experiments
- Conclusion

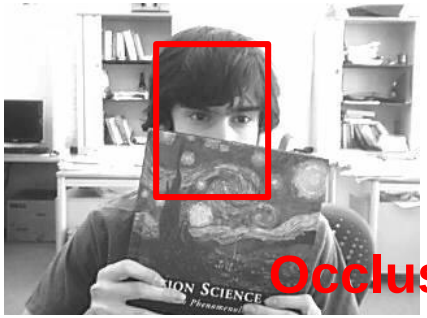
Introduction

- **Problem**: track arbitrary object in video given location in first frame

- **Classic tracking methods**:
  - Discriminative approaches
    - Tracking as a classification problem, where an online binary classifier is trained to separate the target from the surrounding background adaptively.
  - Generative approaches
    - Build a target representation and then searches for the region that is most similar to the target.

Introduction

- Challenges in tracking task



.....

Motivation

- **Dictionary learning (DL)** based object tracking
  - ✓ The aim of DL is to learn a set of basis from the training data in a manner that the dictionary would faithfully represent a given visual signal.
  - ✓ Some DL based trackers have exploited a **binary classification formulation** of the tracking problem in order to improve tracking performance.

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

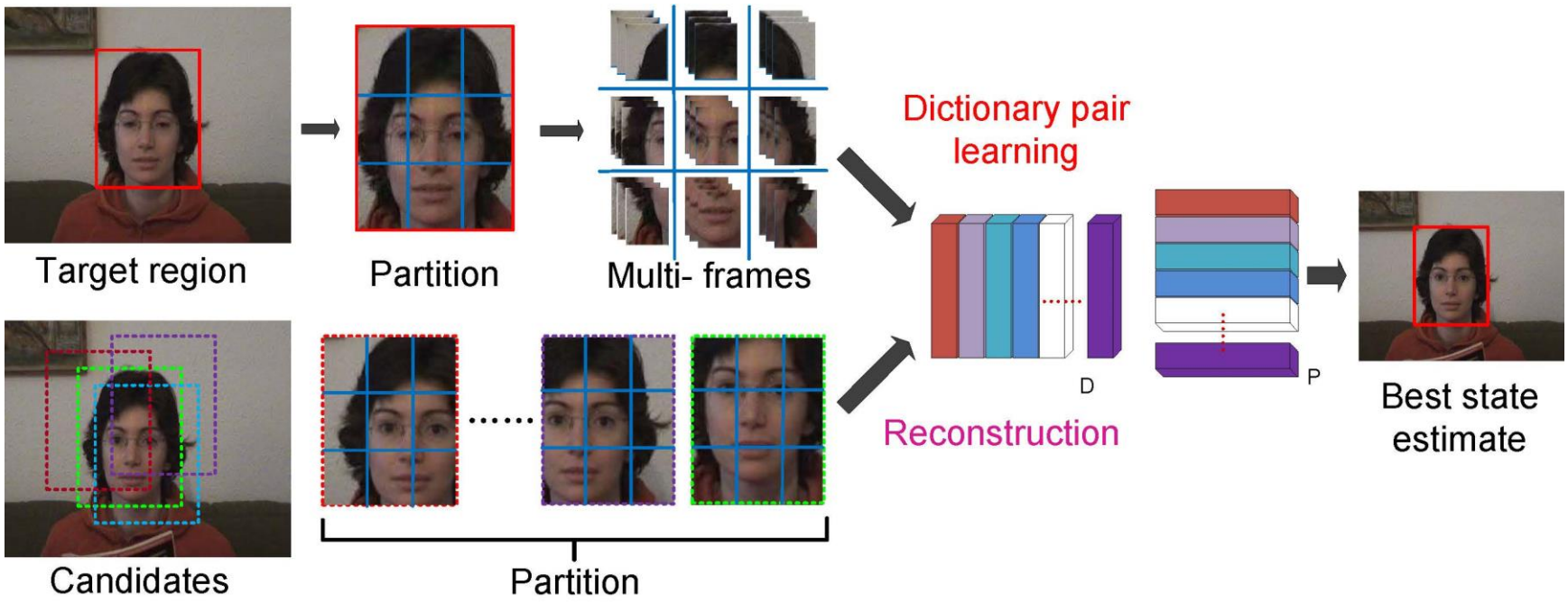
Motivation

- **Dictionary learning (DL)** based object tracking
  - However, some DL based trackers have demonstrated **difficulties in discriminating ambiguities** when the target shares geometric and appearance similarities with the clutter in the background. The limitations mainly include as:
    - Some existing methods select those patches from the target region to represent the positive samples, while patches far away from the target are chosen to represent the negative samples. In this manner, the dictionaries fail to faithfully represent the estimated target candidate, particularly when the target is subjected to changes in pose and appearance.;
    - The coefficients obtained over such dictionaries remain non-discriminative of the target from its corresponding background;
    - Some negative samples thus selected, introduce outliers, which forces that tracker to drift away from the target.

  ✓ Thus, we propose a novel tracking algorithm based on multi-class dictionary learning and collect all training samples from the target region.

Proposed method

- Overall framework



- Object is divided into multiple patches;
- Patches from same locations can be used for training a sub-dictionary;
- Final tracking results can be obtain by fusing reconstruction error from multiple patches.

Proposed method

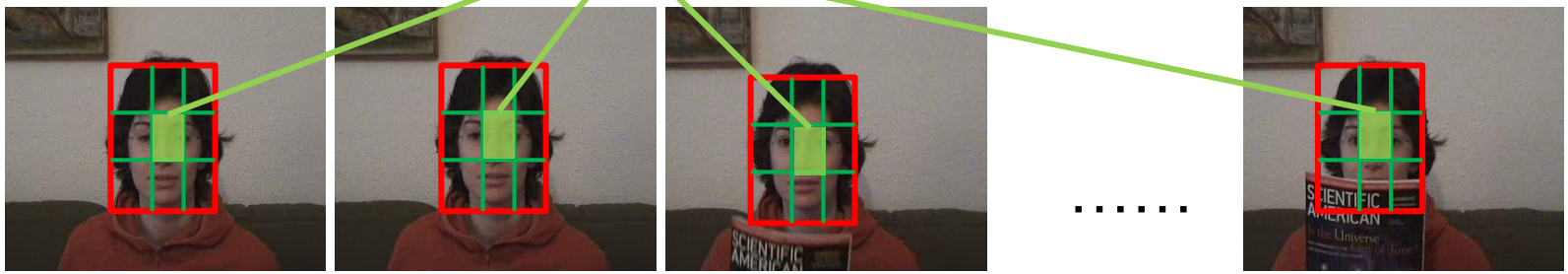
Problem Formulation

✓ Basic dictionary learning model:

$$\underset{\mathbf{D},\mathbf{C}}{\operatorname{argmin}} \sum_{m=1}^{M} \|\mathbf{X}_m - \mathbf{D}_m\mathbf{C}_m\|_F^2 + \lambda\Psi(\mathbf{C}_m)$$

$$\mathbf{X}_m = [X_m^1, X_m^2, \cdots, X_m^N]$$

Collect patches from the *m*-th part of the target



The *m*-th part as an example

- Extract image patches from the *m*-th part of the target, all patches can be collected to form: X_m;
- Use collected patches X_m to learn a sub-dictionary D_m for the *m*-th part of the target.

Proposed method

Problem Formulation

- ✓ Inspired from [1], we introduce an analysis dictionary **P**, then the coefficients **C** can be directly estimated efficiently using **C = PX**. Thus, we have the following objective function:

$$\underset{\mathbf{P},\mathbf{D}}{\operatorname{argmin}} \sum_{m=1}^{M} \|\mathbf{X}_m - \mathbf{D}_m\mathbf{P}_m\mathbf{X}_m\|_F^2 |+ \lambda\|\mathbf{P}_m\mathbf{X}_m^c\|_F^2$$

$$s.t.\ \|\mathbf{d}_i\|_2^2 \leq 1$$

- ✓ Further, we propose Structural Patch-based Dictionary Pair Learning (SPDPL) algorithm:

$$< \mathbf{P}^*, \mathbf{D}^*, \mathbf{C}^* > = \underset{\mathbf{P},\mathbf{D},\mathbf{C}}{\operatorname{argmin}}\{ \sum_{m=1}^{M} \|\mathbf{X}_m - \mathbf{D}_m\mathbf{C}_m\|_F^2 + \lambda_1\|\mathbf{P}_m\mathbf{X}_m - \mathbf{C}_m\|_F^2 + \lambda_2\|\mathbf{P}_m\mathbf{X}_m^c\|_F^2 + \beta\|\mathbf{C}_m - \bar{\mathbf{C}}_m\|_F^2\} \quad s.t.\ \|\mathbf{d}_i\|_2^2 \leq 1$$

- X_m^c denotes the whole training set except X_m, so the term $\|P_m X_m^c\|_F^2$ ensure the sub-dictionary P_m projects a sample from other classes to be the null space, which can improve the dictionary discriminative ability.
- $\bar{C}_m$ denotes the mean of C_m, the term $\|C_m - \bar{C}_m\|_F^2$ is used to reduce the variation of the coding coefficients of each class.

[1] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Projective dictionary pair learning for pattern classification,” in NIPS, 2014, pp.793–801.

Proposed method

- Optimization Solution
  - ✓ The objective function concerning all variables **D**, **P** and **C** has no closed-form solution;
  - ✓ We use an iterative optimization method to update one variable by keeping the other fixed (details can be found in the paper).

Proposed method

- Bayesian theorem based tracking
  - ✓ The posteriori probability can be inferred recursively,

$$p(x_t|y_{1:t}) \propto p(y_t|x_t) \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}$$

  - ✓ The target state can be estimated by the maximum a posteriori estimation,

$$\hat{x}_t = \underset{x_t^i}{\operatorname{argmax}}\, p(x_t^i|y_{1:t})$$

- Observation model
  - ✓ The posteriori probability can be inferred recursively,

$$e_i^j = \left\| y_t^{ij} - D_j P_j y_t^{ij} \right\|_F^2$$

where y_t^{ij} denotes the *j*-th patch in the *i*-th candidate, D_j and P_j denote the dictionary pairs corresponding to the *j*-th patch.

  - ✓ The observation likelihood can be measured using,

$$p(y_t^i|z_t^i) \propto \exp\{-\sum_j^M e_i^j\}$$

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

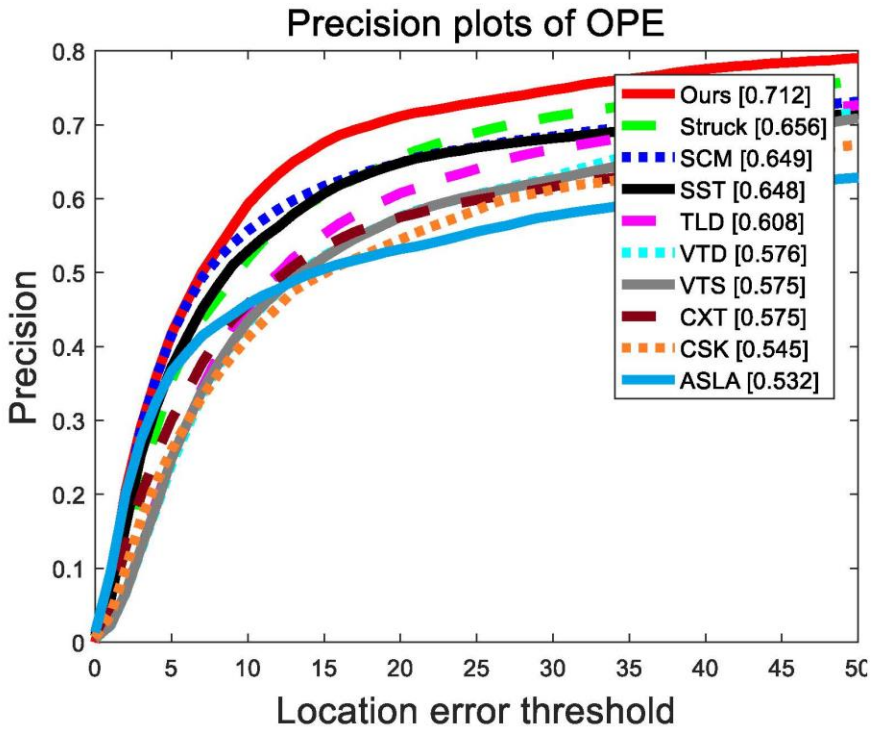
Experiments

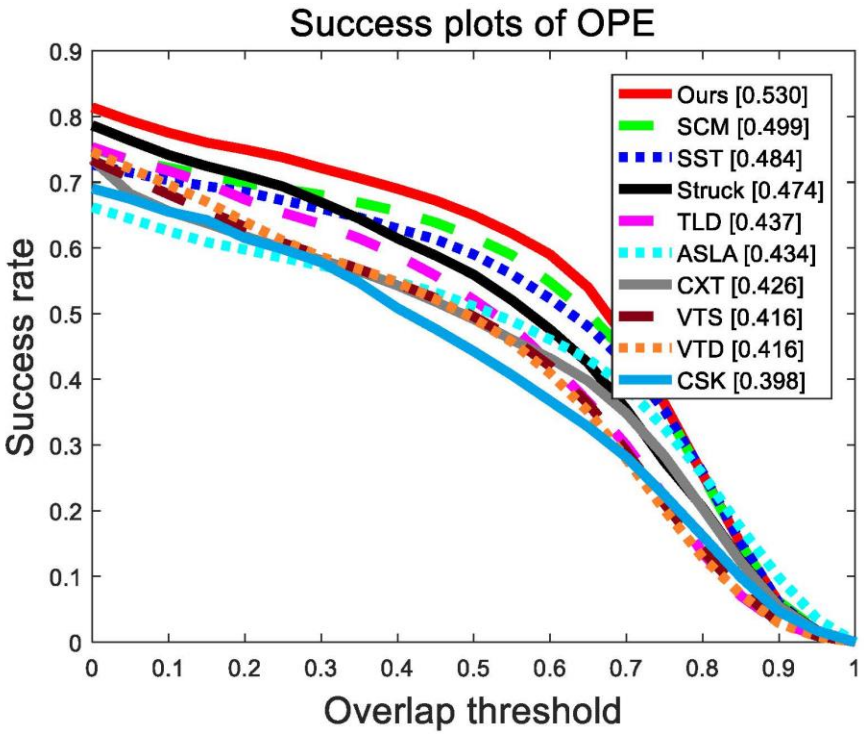
- Experimental setup
  - ✓ Target region is normalized to a 32*32 size;
  - ✓ Each target is divided into M=9 patches;
  - ✓ Evaluate our method on Object Tracker Benchmark;
  - ✓ Add FCT[1] and SST[2] methods into comparison experiments.

[1] K. Zhang, L. Zhang, and M.-H. Yang, “Fast compressive tracking,”IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 10, pp. 2002–2015, 2014.
[2] T. Zhang, S. Liu, C. Xu, S. Yan, B. Ghanem, N. Ahuja, and M.-H. Yang, “Structural sparse tracking,” in CVPR. IEEE, 2015, pp. 150–158.

Experiments

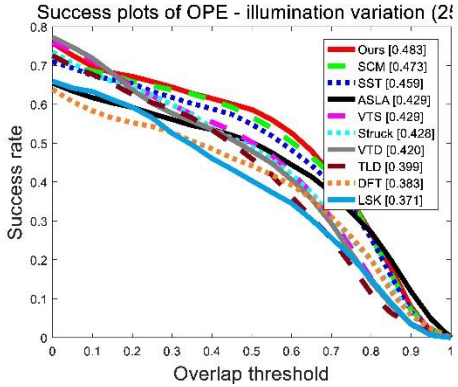
- Experimental results
  - ✓ Overall performance comparison of OPE using precision and success plots.

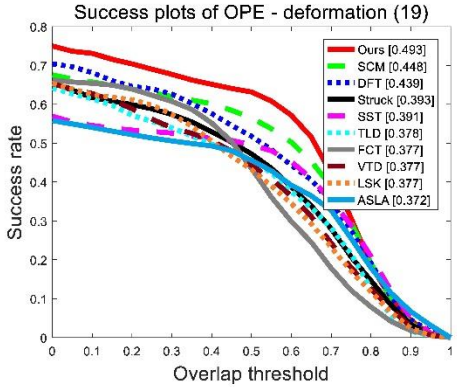


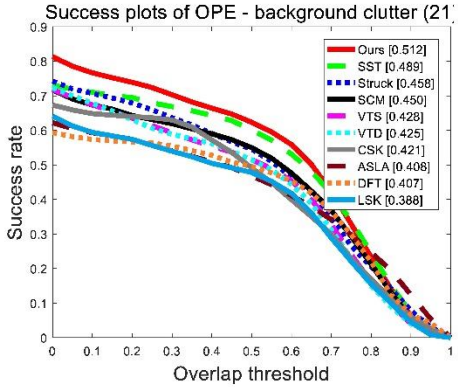


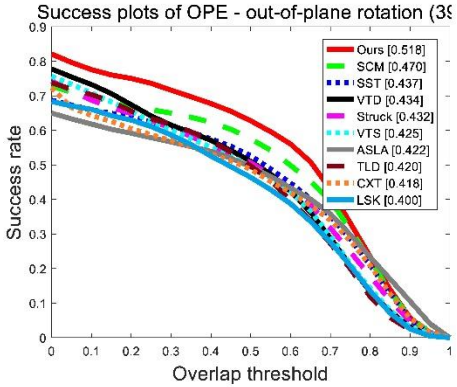
Experiments

- Experimental results
  - ✓ Overlap success plots over four tracking challenges:









Experiments

- Experimental results
  - ✓ Partial Tracking results of SST, Struck, TLD, SCM, IVT, ASLA and our tracker on 4 challenges.

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Experiments

- Experimental results
  - ✓ Frames Per Second (FPS) comparison:

Table 1. FPS comparison of the partial top trackers.

Algorithm	SCM	SST	ASLA	IVT	ℓ_1	Ours
FPS	0.5	2.1	1.1	16.5	0.3	7.5

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

Conclusion

- A novel visual tracking framework based on structural patch-based dictionary pair learning is proposed;
- Tracking is transformed into a multiclass classification problem by regarding all patches from the same part of the target as one class;
- Training set only considers target information, the effect of background can be reduced thus resulting in more accurate tracking performance;
- A simple but effective observation model is designed to obtain an optimal candidate.

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Thanks !