Visual Tracking with Weighted Adaptive Local Sparse Appearance Model via Spatio-Temporal Context Learning

Zhetao Li, *Member, IEEE,* Jie Zhang, Kaihua Zhang, *Member, IEEE,* and Zhiyong Li, *Member, IEEE*

***Abstract*—Sparse representation has been widely exploited to develop an effective appearance model for object tracking due to its well discriminative capability in distinguishing the target from its surrounding background. However, most of these methods only consider either the holistic representation or the local one for each patch with equal importance, and hence may fail when the target suffers from severe occlusion or large-scale pose variation. In this paper, we propose a simple yet effective approach that exploits rich feature information from reliable patches based on weighted local sparse representation that takes into account the importance of each patch. Specifically, we design a reconstruction-error based weight function with the reconstruction error of each patch via sparse coding to measure the patch reliability. Moreover, we explore spatio-temporal context information to enhance the robustness of the appearance model, in which the global temporal context is learned via incremental subspace and sparse representation learning with a novel dynamic template update strategy to update the dictionary, while the local spatial context considers the correlation between the target and its surrounding background via measuring the similarity among their sparse coefficients. Extensive experimental evaluations on two large tracking benchmarks demonstrate favorable performance of the proposed method over some state-of-the-art trackers.**

***Index Terms*—Visual tracking; sparse representation; template update; spatio-temporal context**

I. Introduction

Visual tracking is an important research topic in computer vision with wide applications in various fields, such as self-driving cars, security and surveillance systems, intelligent transportation vision based controls [1]. Visual tracking continually infers the states of an annotated (manually labeled or detected in the first frame) target object in a video sequence. Although visual tracking has long been studied for several decades and much progress has been made in recent years [2]–[18], it remains a challenging task to develop a robust tracking algorithm because the appearance of the tracked target may suffer from severe variations caused by significant pose variation, complicated background clutter, drastic illumination variation, etc.

An effective appearance model plays a key role in ensuring the robustness of a tracking system, thereby attracting much attention in recent years [9], [11]–[25]. Numerous effective representations have been proposed to design the appearance models that can be categorized into either generative [19], [25]–[29] or discriminative models [11]–[18], [22], [30]–[38]. Generative models typically learn an appearance model to represent the target appearance and then use the model to search for the image region with maximal similarity. Generally, the representations for constructing generative appearance models include GMMs [39], color histograms [40], subspace representation [19], and sparse representation [22], [23], [25], [41], [42]. In [39], Jepson et al. proposed a GMM based representation with an online expectation maximization algorithm to overcome target appearance variations during tracking. In [40], Adam et al. utilized a set of local image patch histograms to represent a target object. In [19], Ross et al. proposed an incremental subspace learning method to learn a subspace representation that can adapt to the target appearance changes. In [43], Kwon and Lee decomposed an observation model into multiple basic observation models that are constructed using the sparse principal component analysis. In [24] Wang and Yeung developed a deep learning based tracker that uses stacked de-noising auto-encoder to learn target presentations from a large number of auxiliary images.

Discriminative models cast the tracking problem as a binary classification task, which employ different discriminative features to train a classifier to separate the target from its surrounding background. Avidan [30] first formulated visual tracking as a binary classification problem, which integrated an off-line SVM based classifier into an optical flow based tracker. Collins et al. [44] proposed a feature selection method to learn the most discriminative features online to separate the target object from the background. Grabner et al. [45] proposed an online boosting feature selection method for tracking. Babenko et al. [46] proposed to employ positive and negative bags to learn a multiple instance learning classifier for visual tracking. Kalal et al. [47] improved the binary classifier by considering the structured unlabeled data for visual tracking. In [48], Hare et al. employed an online structured output SVM classifier for robust tracking which can alleviate the effect of wrongly labeling samples. Zhang et al. [12] proposed a multi-expert restoration scheme to address the drift problem in tracking. Recently, Henriques et al. [49]

Zhetao Li is with the College of Information Engineering, Xiangtan University, Hunan 411105, China. E-mail: liztchina@gmail.com. (*corresponding author*)

Jie Zhang is with the College of Information Engineering, Xiangtan University, Hunan 411105, China. E-mail: jiezhang017@gmail.com.

Kaihua Zhang is with the Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China. E-mail: zhkhua@gmail.com.

Zhiyong Li is with the College of Computer Science and Electronic Engineering of Hunan University, Changsha, China, 410082. E-mail: zhiyong.li@hnu.edu.cn

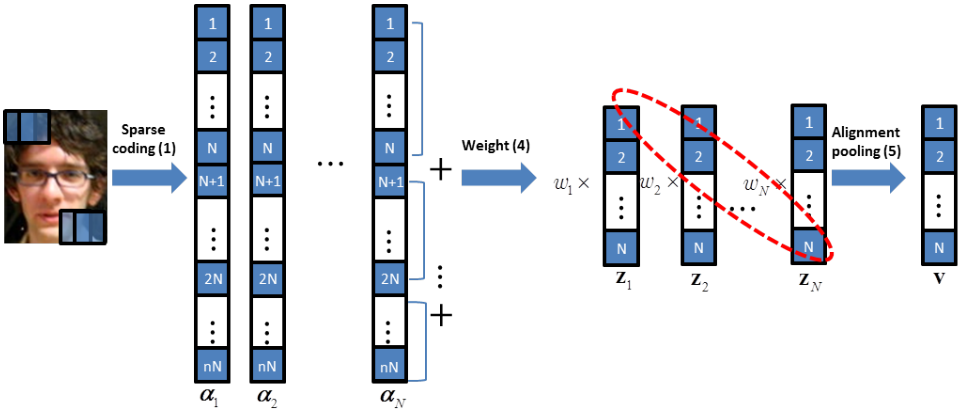


Fig. 1: Flowchart of the proposed weighted local sparse appearance model.

proposed a fast tracker which exploits the circulant structure of the kernel matrix for kernalized correlation filters (KCF) that can be efficiently solved by the fast Fourier transform algorithm. Li and Zhu [14] improved the KCF tracker [49] by integrating a scale adaptive scheme and color-naming features. Ma et al. [13] employed features from a hierarchial layers of convolutional neural networks (CNNs) to learn an effective KCF representation for robust visual tracking.

Recently, sparse representations have been widely exploited in visual tracking [23], [25], [50], [51], which can be categorized into holistic and local sparse representation appearance models. In [25], Mei and Ling exploited a holistic sparse representation of the target appearance for visual tracking that is learned via optimizing an ℓ_1 minimization problem. Li et al. [52] extended this work using the orthogonal matching pursuit algorithm to solve the optimization problem efficiently. Bao et al. [50] further improved the efficiency via the accelerated proximal gradient approach. However, these sparse representation-based trackers take into account the holistic templates of the targets, which are sensitive to severe partial occlusion and pose variations. In [51], Liu et al. proposed a local sparse appearance model that is integrated into the mean shift algorithm to enhance tracking robustness. However, this tracker is based on a static local dictionary obtained from the first frame and has a high probability of failing in dynamic scenes. In [23], Jia et al. presented a local sparse appearance model that employs an alignment-pooling method to combine the histograms of local sparse codings of each patch, in which the dictionary is updated in an online manner to handle target appearance variations, thereby achieving favorable performance on some challenging scenes.

Although demonstrated success of the trackers based local sparse appearance model [23], [25], [50], [51], their performance on the recently tracking benchmark [53] is not favorable. For example, as reported by the benchmark, the AUC score of success plots of OPE for the ASLA tracker [23] is just 0.434, which is much lower than the KCF based trackers, e.g., 0.514 for KCF [49] and 0.567 for SAMF [14] as reported in [14]. We note that the performance of ASLA tracker still has plenty of room for improvement if we take into account the importance of different local patches and attempt to integrate the spatio-temporal context information. In this paper, we propose a simple yet effective method by combining the weighted local sparse model and spatio-temporal context information. The proposed method is motivated by the ASLA tracker [23], but takes into account the patch importance to measure the reliability of each patch with a reconstruction-error-based weight function of reconstruction error. Furthermore, we employ the spatio-temporal context information, which incorporates the global temporal context via an incremental subspace and sparse representation learning with a novel online template update strategy and the local spatial context by taking into account the correlation between the target and its surrounding background.

The contributions of this work are summarized as follows:

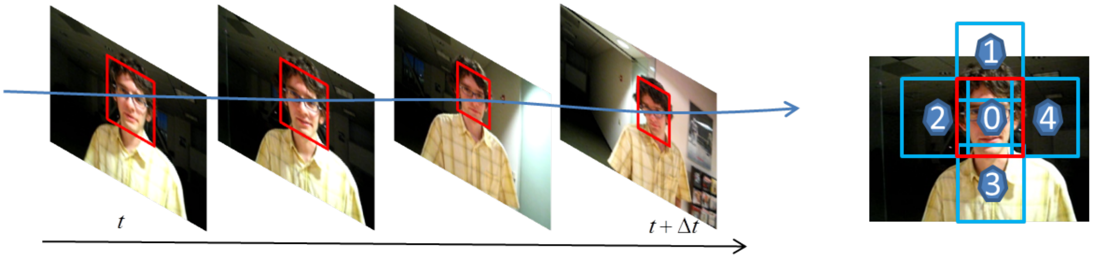
1) We propose a weighted local sparse model that takes into account the reliability of each local patch and the spatio-temporal context information, thereby well adapting to target appearance variations in challenging cases.
2) We employ an adaptive template update strategy that combines the incremental subspace learning and sparse representation to update the dictionary with dynamic templates, which makes our tracker effectively deal with partial occlusion and the drifting problem.
3) Experiments on a large-scale tracking benchmark demonstrate that the proposed tracker performs favorably against several state-of-the-art methods.

II. Methodology

In this section, we first show how the local appearance model of the target is generated by weighted local sparse coding method. Next, we describe how to integrate spatio-temporal context information for tracking. Finally, we present the proposed algorithm in detail.

A. Weighted Local Sparse Appearance Model

In this work, we employ a local sparse representation with a set of sparse coefficients to model the appearance of target patches. Given a set of target templates $\mathrm{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n]$, where $\mathbf{t}_i$ denotes the image intensity vector of the tracked target, we sample a set of overlapped local image patches



(a) Global Temporal Context

(b) Local Spatial Context

Fig. 2: Illustration of temporal and spatial context constraints.

inside the target region by sliding a window with a fixed size. These local patches are used to construct the dictionary that encodes the local pathes inside the possible candidate regions, i.e., $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{n\times N}] \in \mathcal{R}^{d\times(nN)}$, where d denotes the dimension of local patch vector, n is the number of target templates and N is the number of sampled local patches inside the target region. Each atom $\mathbf{d}_i$ in the dictionary $\mathbf{D}$ is achieved by ℓ_2 normalization on the local image patch vector. Since each atom in $\mathbf{D}$ represents a fixed part of the target object, $\mathbf{D}$ collects all the structure information of the templates, which shares the commonality of different templates, thereby being robust to template variations.

Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathcal{R}^{d\times N}$ denote the candidate sample representation that we extract local patches inside it and transform them into vectors in the same way. With the overcomplete dictionary $\mathbf{D}$ and sparsity assumption, each local patch $\mathbf{y}_i$ can be represented by a linear combination of only a few atoms in the dictionary by solving

$$\min_{\boldsymbol{\alpha}_i} = \|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_1, s.t., \boldsymbol{\alpha}_i \succeq 0, \quad (1)$$

where $\boldsymbol{\alpha}_i \in \mathcal{R}^{nN}$ is the sparse coefficient vector corresponding to the local image patch $\mathbf{y}_i$. Here, solving (1) needs to tune the parameter λ, which is nontrivial. Fortunately, we can employ the simplex and sparse representation method proposed by [54] to modify the constraint in (1) into the ℓ_1 ball constraint as $\boldsymbol{\alpha}_i \succeq 0$ and $\mathbf{1}^\top\boldsymbol{\alpha}_i = 1$, and hence we do not need to tune λ anymore.

The sparse coefficient vector $\boldsymbol{\alpha}_i$ is divided into several segments according to their corresponding templates, i.e., $\boldsymbol{\alpha}_i = [\boldsymbol{\alpha}_i^{(1)\top}, \boldsymbol{\alpha}_i^{(2)\top}, \ldots, \boldsymbol{\alpha}_i^{(n)\top}]^\top$. In [23], all the segment coefficient vectors are equally weighted to yield the representation $\mathbf{z}_i$ for the patch $\mathbf{y}_i$

$$\mathbf{z}_i = \frac{1}{C}\sum_{k=1}^{n}\boldsymbol{\alpha}_i^{(k)}, \quad (2)$$

where C is a normalization constant. Since the template set T contains the target object with some appearance variations, the patches with less reconstruction errors with their sparse codes in these templates should be weighted more than others for more robust representation. To this end, for the patch $\mathbf{y}_i$, we design a simple weight function with respect to its reconstruction error

$$w_i = \frac{e^{-\frac{\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2}{\sigma}}}{\sum_{i=1}^{N} e^{-\frac{\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2}{\sigma}}}, \quad (3)$$

where σ is a constant that balances the patch weight. Coincidentally, (3) can be directly derived from solving $\min_{w_i}(\sum_i w_i\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \sigma w_i \log w_i)$, which is similar to the objective function of the agglomerative fuzzy K-Means algorithm [55], where the left term is similar to the standard K-Means algorithm with predefined center $\mathbf{D}\boldsymbol{\alpha}_i$, and the right term is added to maximize the negative objects-to-clusters membership entropy in the clustering process, which can simultaneously minimize the within cluster dispersion and maximize the negative weight entropy to determine clusters to contribute to the association of objects [56].

Then, the representation $\mathbf{z}_i$ in (2) is reformulated as

$$\mathbf{z}_i = w_i\sum_{k=1}^{n}\boldsymbol{\alpha}_i^{(k)}, i = 1, \ldots, N, \quad (4)$$

which differentially accounts for the reliability of each local patch in the templates. All the vectors $\mathbf{z}_i$ of the local patches in a candidate region construct a square matrix $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N]$. Then, as [23], we employ an alignment-pooling strategy to take the diagonal elements of the square matrix $\mathbf{Z}$ as the pooled feature

$$\mathbf{v} = \mathrm{diag}(\mathbf{Z}), \quad (5)$$

where $\mathbf{v}$ denotes the pooled feature vector, which integrates the appearance and spatial information from the candidate target region, thereby encoding the structure information of the target.

Figure 1 illustrates the flowchart of our proposed weighted local sparse appearance model. Different from Jia et al.'s model [23] that utilizes the same weight for each patch, the proposed method weights the contaminated local patch less while the reliable patch more, thereby mitigating the side effect of noise introduced by target appearance variations.

B. Spatio-Temporal Context Integration

The temporal and spatial context information is important for robust tracking [57]. The target appearance changes gradually between two consecutive frames due to the high frame rate of current video sequences (about 25 frames per second), and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2018.2839916, IEEE Transactions on Image Processing

all of the historical appearance variations have some influences on the current appearance state estimation. Meanwhile, since the target moves smoothly from one location to another location, the spatial context presents strong correlation between the target and its surrounding backgrounds. In this section, we introduce how to incorporate the spatio-temporal context information into our appearance model for robust tracking.

1) Global Temporal Context: As shown by Figure 2, we collect the tracking results of the target object and then carry out the incremental subspace learning method as [19], which not only preserves the collected common observations, but also well adapts to the appearance variations. We model the estimated target by a linear combination of the PCA basis vectors and additional trivial templates [25]

$$\mathbf{p} = \mathbf{U}\mathbf{q} + \mathbf{e} = [\mathbf{U}\ \mathbf{I}][\mathbf{q}\ \mathbf{e}]^\top, \tag{6}$$

where $\mathbf{p}$ represents the estimated target vector, $\mathbf{U}$ is the matrix composed of eigenbasis vectors, $\mathbf{q}$ is the coefficients of the eigenbasis vectors and $\mathbf{e}$ represents the pixels in $\mathbf{p}$ that are corrupted. In [23], the sparsity constraints are enforced on the estimation of both $\mathbf{q}$ and $\mathbf{e}$. However, enforcing sparse constraint on the coefficients $\mathbf{q}$ may result in losing useful information for tracking because of the orthogonality of the PCA basis matrix $\mathbf{U}$. Therefore, we propose a new template update formula by only enforcing sparse constraint on $\mathbf{e}$ because the error caused by occlusion or noise often owns sparse distribution

$$\min_{\mathbf{q},\mathbf{e}} \|\mathbf{p} - \mathbf{U}\mathbf{q} - \mathbf{e}\|_2^2 + \lambda\|\mathbf{e}\|_1. \tag{7}$$

To solve (7), we first fix $\mathbf{e}$, and optimize $\mathbf{q}$ as

$$\mathbf{p} \leftarrow \mathbf{U}^\top(\mathbf{p} - \mathbf{e}), \tag{8}$$

and then, we fix $\mathbf{p}$, and optimize $\mathbf{e}$ as [17]

$$\mathbf{e} \leftarrow \mathrm{sign}(\mathbf{p} - \mathbf{U}\mathbf{q}) \max(0, \mathrm{abs}(\mathbf{p} - \mathbf{U}\mathbf{q}) - \lambda). \tag{9}$$

We iteratively update (8) and (9) until convergence. Moreover, the PCA basis matrix $\mathbf{U}$ is incrementally updated by the incremental subspace learning method [19]. After obtaining the solution $\mathbf{q}$ of (7), we have the representation of $\mathbf{p}$ as $\tilde{\mathbf{p}} = \mathbf{U}\mathbf{q}$. Then, we measure the similarity between the reconstructed representation $\tilde{\mathbf{p}}$ and the template in the template set T as

$$\rho_i = \frac{\tilde{\mathbf{p}}^\top \mathbf{t}_i}{\|\tilde{\mathbf{p}}\|_2^1 \|\mathbf{t}_i\|_2^1}, i = 1, \dots, n. \tag{10}$$

If $\rho_i < 0.65, i = 1, \dots, n$, which means that the new tracking result has small similarity with the template set, we don't update the set T. Otherwise, if $\rho_i > 0.85$, which means that the current tracking result can be well represented by the template set, so we don't need to update the template set either. Only when $0.65 \leq \rho_i \leq 0.85, i = 1, \dots, n$, we use $\tilde{\mathbf{p}}$ to update the template set T by replacing one element therein. The main steps of the template update algorithm are summarized in Algorithm 1.

Algorithm 1 Template Update

```
Input: Observation vector of the target estimation $\mathbf{p}$, PCA eigenbasis vector matrix $\mathbf{U}$, template set $\mathrm{T}$
  1) Solve (7) to yield $\mathbf{q}$ and $\mathbf{e}$, and compute the reconstruction representation $\tilde{\mathbf{p}} = \mathbf{U}\mathbf{q}$ ;
  2) Compute $\rho_i$ via (10);
  3) if $0.65 \leq \rho_i \leq 0.85$
  4)     Generate a sequence of numbers in an ascending order and normalize them into $[0, 1]$ as the probability for template update, and generate a random number between 0 and 1 to select which template to be discarded, and add $\tilde{\mathbf{p}}$ to the end of the template set $\mathrm{T}$;
  5) end if
Output: New template set $\mathrm{T}$.
```

2) Local Spatial Context: As shown by Figure 2 (b), the local spatial context information is derived from the regions surrounding the target object (here, we utilize five surrounding patches including the target patch as local context information). The works in [57], [58] show that the local context information including supporters and distracters enhances the robustness of the tracker, even when the target is partially occluded. However, different from [58] that constructs complex relative motion model between the target and auxiliary objects and [57] that employs boosting method to construct a strong classifiers to combine the weak correlation between every contributor and the target, we propose a simple yet effective method that explores the similarity between the target appearance feature $\mathbf{v}_{t-1}$ computed via (5) in the current frame and the i-th candidate context region appearance feature set $\{\mathbf{v}_t^{i,k}\}_{k=0}^4$ to account for the local spatial context constraints

$$f_t^i = e^{-(\|\mathbf{v}_{t-1} - \mathbf{v}_t^{i,0}\|_2^1 - \sum_{k=1}^4 \|\mathbf{v}_{t-1} - \mathbf{v}_t^{i,k}\|_2^1)}. \tag{11}$$

(11) measures the weak correlation between the target appearance and its local surrounding context region appearances, which yields a high value if increasing the similarity between the target and its center context region while reducing the similarity with its surrounding context regions, thereby encoding the structure context information to help discriminate target from background.

C. Proposed Tracking Algorithm

The proposed tracking algorithm is formulated under a particle filtering framework. Given the observation set $o_{1:t} = \{o_1, \dots, o_t\}$ up to frame t, the target state variable s_t can be computed via maximizing a posteriori estimation

$$\hat{s}_t = \arg\max_{s_t^i} p(s_t^i | o_{1:t}), \tag{12}$$

where s_t^i represents the state of the i-th sample. The posterior probability $p(s_t|o_{1:t})$ can be recursively inferred by the Bayesian theorem

$$p(s_t|o_{1:t}) \propto p(o_t|s_t) \int p(s_t|s_{t-1}) p(s_{t-1}|o_{1:t-1}) ds_{t-1}, \tag{13}$$

Algorithm 2 WALSA based Tracking

```
Input: Target state $\hat{s}_{t-1}$, target appearance feature $\mathbf{v}_{t-1}$, template set $\mathrm{T}$, the PCA basis matrix $\mathbf{U}$;
1) Sample $m$ candidate particles $\{s_t^i\}_{i=1}^m$ with the motion model $p(s_t^i|\hat{s}_{t-1})$;
2) For each particle $s_t^i$, compute its observation model $p(o_t|s_t^i)$ by (14);
3) Estimate the optimal state $\hat{s}_t$ by (12), and get its corresponding reconstruction representation $\tilde{\mathbf{p}}$ via solving (7), and compute $\rho_i$ via (10);
4) if $0.65 \leq \rho_i \leq 0.85$
5)     Update the template set $\mathrm{T}$ via Algorithm 1;
6)     Update the target appearance feature $\mathbf{v}_t$ with the new template set $\mathrm{T}$
7)     Update the PCA basis matrix $\mathbf{U}$;
8) else
9)     $\mathbf{v}_t \leftarrow \mathbf{v}_{t-1}$
10) end if
Output: Target state $\hat{\mathbf{s}}_t$, updated target appearance feature $\mathbf{v}_t$, updated template set $\mathrm{T}$, updated PCA basis matrix $\mathbf{U}$.
```

where $p(s_t|s_{t-1})$ represents the dynamic model while $p(o_t|s_t)$ is the observation model. We assume that the target state parameters are independent as scalar Gaussian distributions and model the motion as Brownian motion [19], i.e., $p(s_t|s_{t-1}) = \mathcal{N}(s_t|s_{t-1}, \Sigma)$, where $\Sigma = \mathrm{diag}(\sigma_x, \sigma_y, \sigma_s)$. In tracking, the posterior probability $p(s_t|o_{1:t})$ in (13) is approximated by a particle filter that m particles $\{s_t^i\}_{i=1}^m$ are sampled with corresponding importance weights $\{\pi_t^i \propto p(o_t|s_t^i)\}_{i=1}^m$, where $p(o_t|s_t^i)$ is the observation model that denotes the likelihood of the observation o_t at state s_t^i that plays an important role in visual tracking. In our method, we formulate the observation model as

$$p(o_t|s_t^i) \propto f_t^i, \tag{14}$$

where f_t^i is computed by (11). The main steps of the proposed algorithm are summarized in Algorithm 2.

III. Experiments

A. Experimental Setup

Our algorithm is implemented in MATLAB that runs at 5 frames per second on an Intel Core i7 CPU machine with 8 G RAM. The regularization parameter for λ is set to 0.01. We set the variances of the affine parameters in the dynamic model to $\{\sigma_x, \sigma_y, \sigma_s\} = \{6, 6, 0.01\}$, and the number of samples for particle filter is set to 600. We manually label the location of the target object in the first frame for each sequence, and normalize each target image patch to 32×32 pixels and extract 16×16 local patches overlapped within the target region with 8 pixels as the sliding step. The parameters n and N for the local appearance model are set to 10 and 9, respectively. We utilize 20 eigenvectors for incremental subspace learning. All the parameters are fixed for all experiments.

B. Evaluation Metrics

We evaluate the proposed WALSA algorithm on two large tracking benchmark datasets [53], [59], termed as OTB50 and OTB100, which contains 50 and 100 fully-annotated video squences. To better evaluate the effectiveness of the proposed tracker, we further add six most recent trackers including DLT [24], KCF [49], DSST [60], TGPR [11], MEEM [12], CNT [17] besides the 29 trackers within the benchmark. For detailed analysis, the videos are categorized into 11 attributes based on different challenging factors including out-of-plane rotation (OPR), scale variation (SV), occlusion (OCC), low resolution (LR), in-plane rotation (IPR), deformation (DEF), background clutters (BC), illumination variation (IV), motion blur (MB), fast motion (FM), and out-of-view (OV).

We employ the success plot and precision plot for quantitative evaluations, in which the success plot is defined on the overlap ratio, i.e., $S = area(B_T \bigcap B_G)/area(B_T \bigcup B_G)$ with the tracked bounding box B_T and the ground truth B_G. The area under curve (AUC) of each success plot is used to rank the evaluated trackers. Meanwhile, the precision plot demonstrates the percentage of frames whose tracked locations are less than 20 pixels to the ground truth. We employ the success plot and precision plot to indicate the one-pass evaluation (OPE), and report the results of OPE for each evaluated tracker.

We have observed from the experiments that the rankings of different trackers on OTB50 and OTB100 are almost the same, and hence in the following sections, for presentation clarity, we only report the overall performance of the top 5 trackers on OTB100, and analyze the results on OTB50.

C. Quantitative Evaluations

1) Overall Performance: Figure 3 shows the over all performance of the top 5 trackers on OTB100, and Figure 4 illustrates the overall performance of the top 10 trackers in terms of success and precision plots, where the proposed WALSA ranks first in terms of success rate and second based on precision rate. In the success plot, the WALSA achieves the AUC score of 0.580 that outperforms the runner-up MEEM by 1.4%. Moreover, the WALSA outperforms the baseline ASLA method [23] by a large margin (0.580 vs. 0.434), which validates the effectiveness of the proposed weighted strategy. Meanwhile, as shown by the precision score in the precision plot, the WALSA algorithm obtains 0.794 which is competitive to the top ranker MEEM (0.830), but significantly outperforms ASLA by 29.8%. In addition, WALSA outperforms the CNN-based trackers CNT [17] and DLT [24], which shows the strong discriminative capability of the features extracted via the weighted local sparse representation learning.

2) Attribute-Based Performance: We further compare the trackers with 11 attributes to clearly analyze the strength and weakness of the proposed algorithm. Figure 5 and Figure 6 show the success plots and precision plots with different attributes, respectively. Among them, the proposed WALSA tracker ranks within top 3 on all attributes in terms of success rate and on 10 out of 11 attributes in terms of precision rate. Specifically, as shown in the success plots, for the videos with attributes of *scale variation*, *occlusion*, *illumination variation*, *motion blur*, *fast motion*, and *out of view*, WALSA achieves the top performance among all the evaluated trackers. For

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2018.2839916, IEEE Transactions on Image Processing

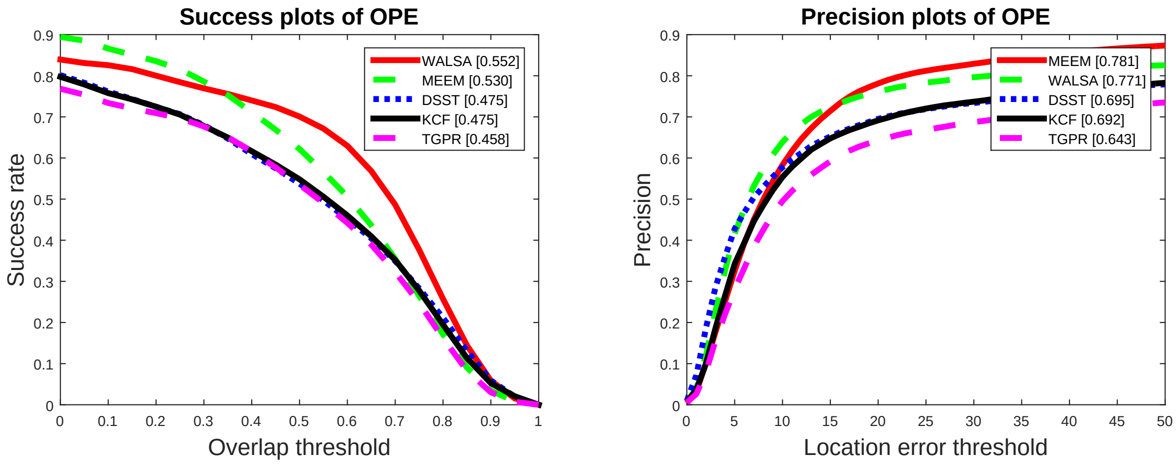


Fig. 3: Success plots and precision plots of OPE for the top 5 trackers on OTB100. The performance score for each tracker is the AUC value that is shown in the legend. Meanwhile, the performance score of precession plot is at error threshold of 20 pixels. Best viewed on color display.

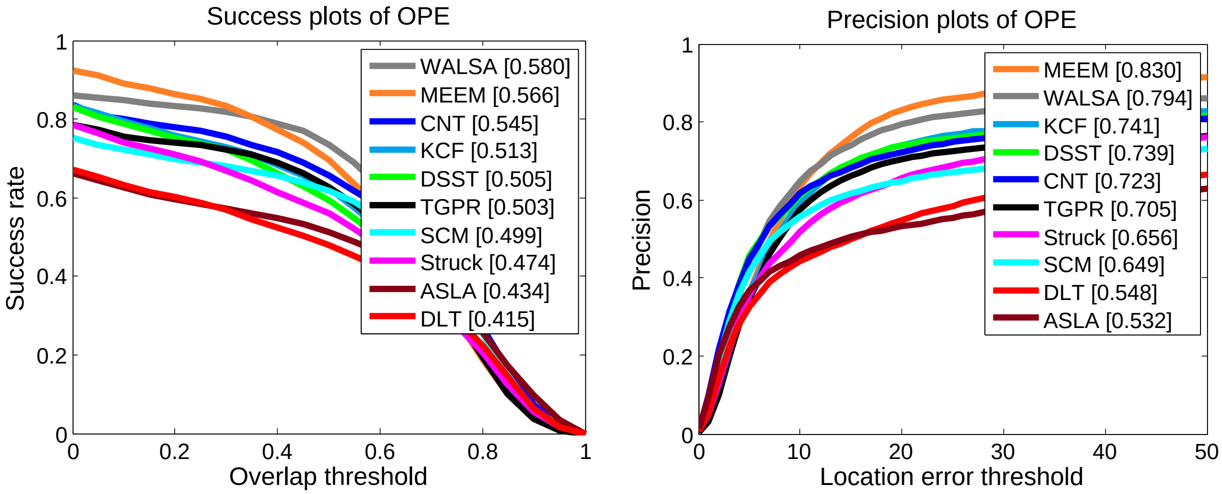


Fig. 4: Success plots and precision plots of OPE for the top 10 trackers on OTB50. The performance score for each tracker is the AUC value that is shown in the legend. Meanwhile, the performance score of precession plot is at error threshold of 20 pixels. Best viewed on color display.

the videos with attributes of *low resolution*, *out-of-plane*, *deformation*, *background clutter*, WALSA ranks second among all trackers, and WALSA ranks third on the videos with the attribute of *in-plane rotation*. Meanwhile, WALSA significantly outperforms ASLA on all attributes in terms of both success and precision rates.

D. Qualitative Comparisons

1) Scale Variation: Figure 7 shows some tracking results in three challenging videos in which the targets suffer from significant scale variations. The person in the *david* sequence moves from a dark room to a bright area that causes his appearance varies much due to illumination changes, pose variations, and a large scale variation with respect to the camera. The STRUCK, DLT and KCF undergo drift to background (e.g., #580, #680, #758). In the *freeman3* video, a person appearance varies much because of the pose variation and low resolution. Moreover, the person suffers from large-scale variation when he moves to the camera. The KCF, STRUCK, DLT, TGPR, DSST, and MEEM drift away to the background (e.g., #299, #360, #382), while the WALSA, CNT and ASLA perform well on all frames. The person in the *singer1* sequence moves far way from the camera with a large-scale variation. The KCF, TGPR, STRUCK, MEEM, and DSST works poorly while the WALSA, CNT, ASLA perform well. The WALSA deals scale variation well due to its adopted alignment-pooling strategy.

2) Deformation: Figure 8 shows the results in three challenging sequences that the targets suffer from severe deformation. The target in the *singer2* sequence suffers from illumination variations and significant deformation. Only the WALSA, CNT, TGPR and KCF perform well at all frames. In the *bolt* sequence, several persons simultaneously appear in the scene that undergo rapid appearance variations because of shape deformation and fast motion. Only the WALSA, CNT

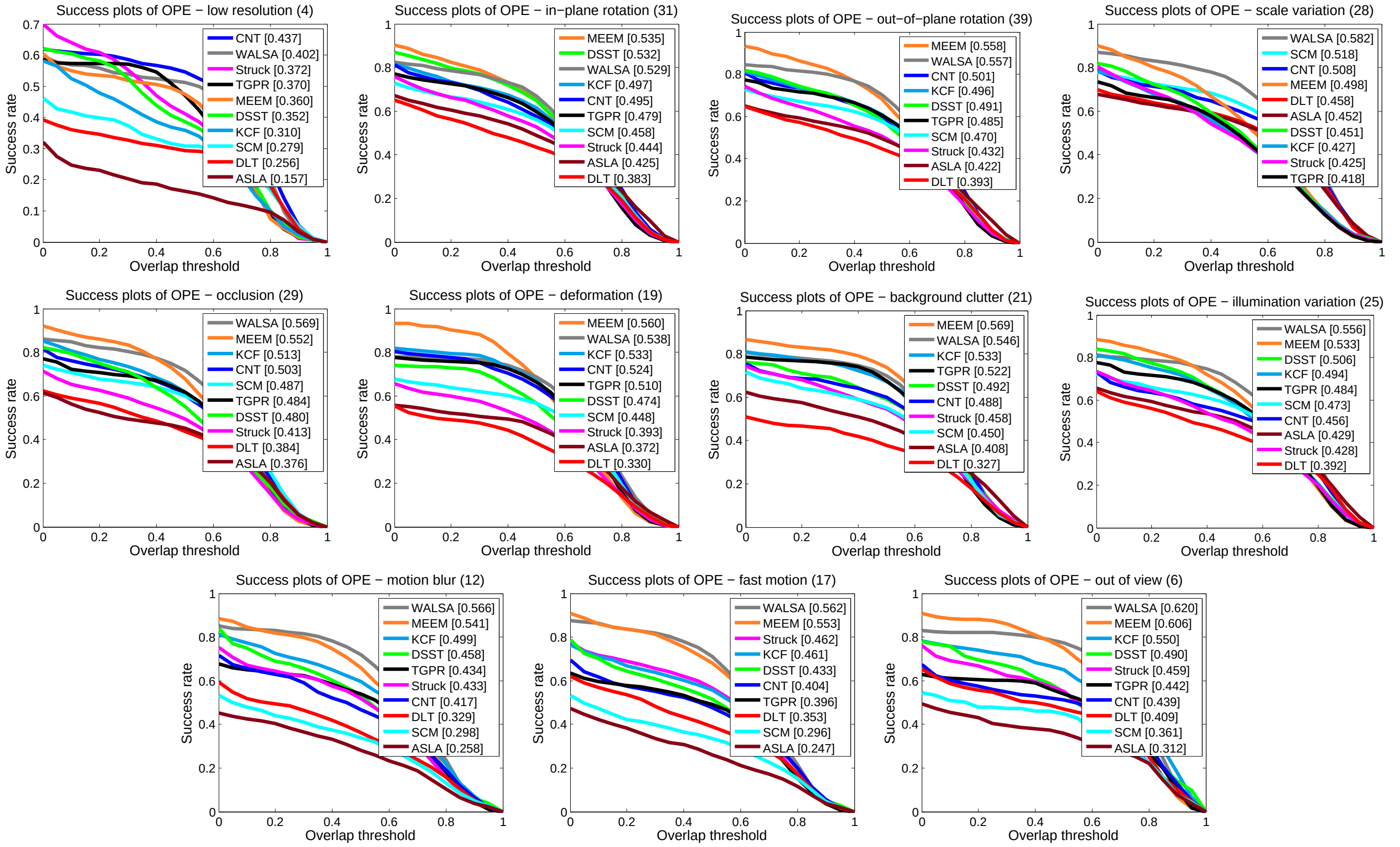


Fig. 5: Success plots of videos with different attributes on OTB50. Best viewed on color display.

and KCF enable to track the targets stably well. The SCM, ASLA, TGPR, STRUCK and SCM drift to the background at the beginning frames (e.g., #15, #55). In the *david3* video, the target has large appearance variations because of non-rigid body deformation. Besides, the target appearance changes much when the person turns around. The SCM and STRUCK lose tracking the target after frame #140. The ASLA and DSST drift to the background when the target turns around. Only the WALSA, CNT, TGPR and KCF perform well on all frames.

3) Occlusion: Figure 9 shows some sampled tracking results of three sequences with target having heavy occlusions. The target in the *jogging-1* sequence is heavily occluded by the lamp post (e.g., #76). Only the WALSA, CNT and MEEM enable to re-detect the target when it reappears in the screen (e.g., #81, #151). In the *suv* sequence, the vehicle undergoes heavy occlusion from dense tree branches (e.g., #515m #550, #680). The TGPR, STRUCK, MEEM and KCF cannot perform well (e.g., #680). In the *woman* sequence, although the target is occluded for a long duration (e.g., #115, #295), the WALSA performs well at most frames, whereas the ASLA perform poorly on this sequence (e.g., #145).

4) Illumination Changes: Figure 10 demonstrates some sampled tracking results in three videos including targets undergoing large illumination changes. A moving car in the *car4* sequence passes underneath a bridge, thereby causing drastic illumination changes. The WALSA performs well despite the large illumination changes at frames #240, #405. The MEEM and KCF suffer from some drift when drastic illumination variation occurs as illustrated by frame #240. In the *skating1* sequence, the target undergoes drastic light changes and rapid pose variations (e.g., #175, #205, #335). The CNT, STRUCK, and WALSA performs consistently well from the beginning frame to the end. In the *trellis* sequence, the target appearance has significant variations in pose. The DLT and ASLA drift away to the background at frames #305, #405. The WALSA, CNT, and STRUCK enable to stably track the target with much more better accuracy than the TGPR, KCF, DSST and MEEM methods.

TABLE I: Effect of setting different values of parameter λ in (7).

λ	0.005	0.008	0.01	0.05	0.1
AUC score	0.562	0.56	0.58	0.575	0.578

E. Ablative Study

We propose three variants of WALSA to further validate the effectiveness of its key components: one only utilizes the weighted local appearance model without spatio-temporal context, one does not include the temporal context, and another one does not include the spatial context. We take the ASLA as the baseline tracker. Figure 11 demonstrates the results of OPE on the benchmark dataset. The results show that without considering spatio-temporal context, the AUC score of WALSA reduces by 12.1%. Meanwhile, the

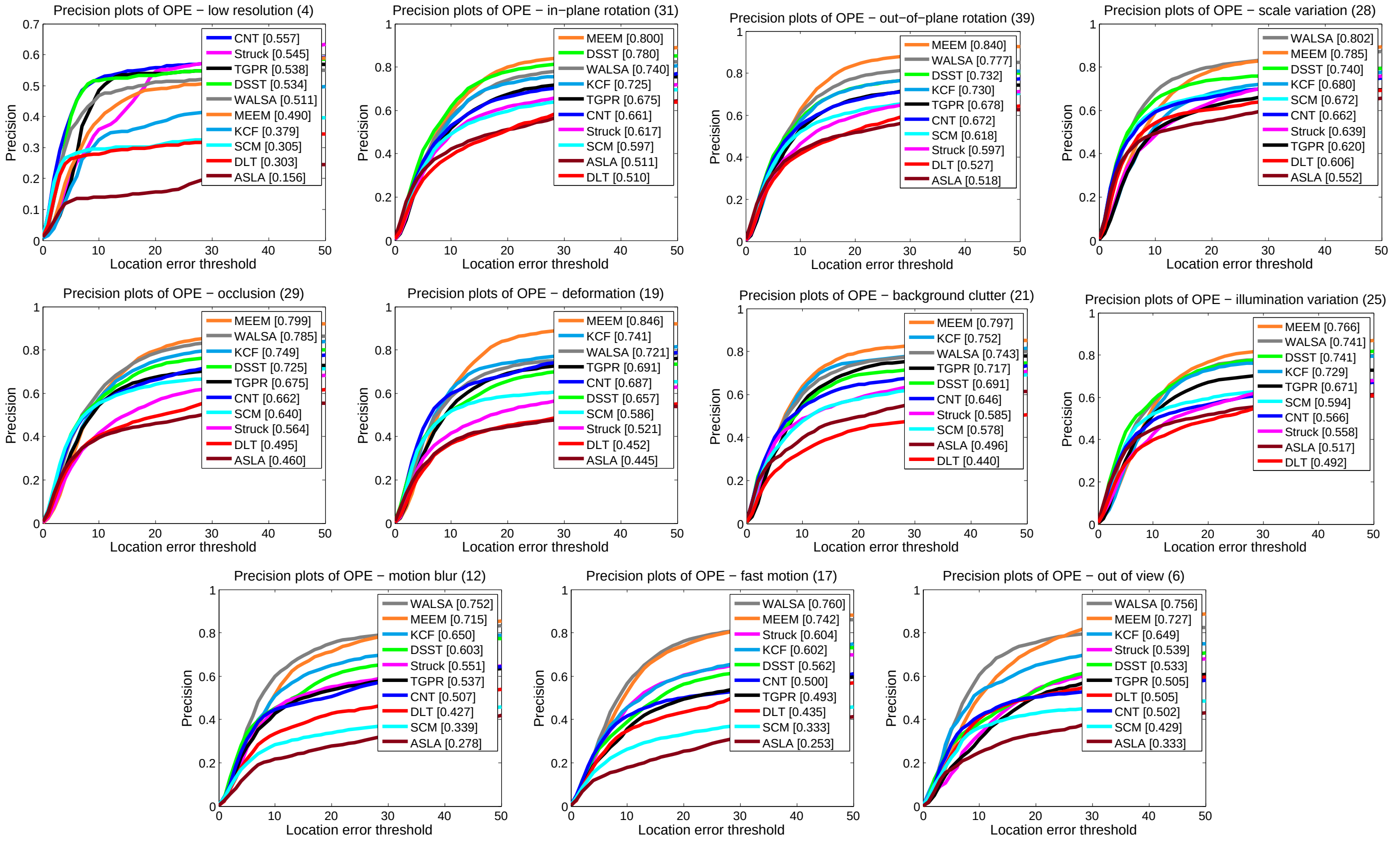


Fig. 6: Precision plots of videos with different attributes on OTB50. Best viewed on color display.

WALSA without spatial context reduces the AUC score by 6.6% while by 4.1% without temporal context, which shows that the temporal context is more important than the spatial context. In addition, even without spatio-temporal context, the WALSA outperforms the ASLA by 2.5%, which validates the effectiveness of the weighted strategy in WALSA.

Table I lists the results of setting different values of parameter λ in (7). We set $\lambda = 0.005, 0.008, 0.01, 0.05, 0.1$, and report the AUC score of success plot for each setting. We observe that the proposed method is not sensitive to this parameter and achieves best performance when $\lambda = 0.01$.

IV. Conclusions

In this paper, we have presented a simple yet effective approach that explores rich feature information from reliable patches based on the weighted local sparse representation. Specifically, we designed a weight function with the reconstruction error of each patch via sparse coding to measure the patch reliability. Moreover, we explored the spatio-temporal context information to enhance the robustness of the appearance model, in which the global temporal context is learned via incremental subspace and sparse representation learning with a novel dynamic template update strategy, while the local spatial context considers the correlation between the target and its surrounding background via measuring the similarity among their sparse coefficients. Extensive experimental evaluations on the large tracking benchmarks demonstrated favorable performance of the proposed method over some state-of-the-art trackers.

Acknowledgements

Zhetao Li is supported in part by National Key Research and Development Program of China under Grant 2018YFB1003702, in part by Hunan Provincial Natural Science Foundation of China for Distinguished Young Scholars under Grant 2018JJ025. Kaihua Zhang is supported in part by NSF of Jiangsu province under grand BK20151529 and under grand BK20170040, in part by and Six talent peaks project in Jiangsu Province under Grant R2017L07, in part by the Natural Science Foundation of China (NSFC) under Grant 61773002, in part by Applied basic research project in Shanxi Province under Grant 201601D011007. Zhiyong Li is partially supported by the National Natural Science Foundation of China (No.61672215, No.91320103), the Special Project on the Integration of Industry, Education and Research of Guangdong Province, China (No.2012A090300003) and the Science and Technology Planning Project of Guangdong Province, China (No.2013B090700003).

References

[1] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel, "A survey of appearance models in visual object tracking," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 4, p. 58, 2013. 1

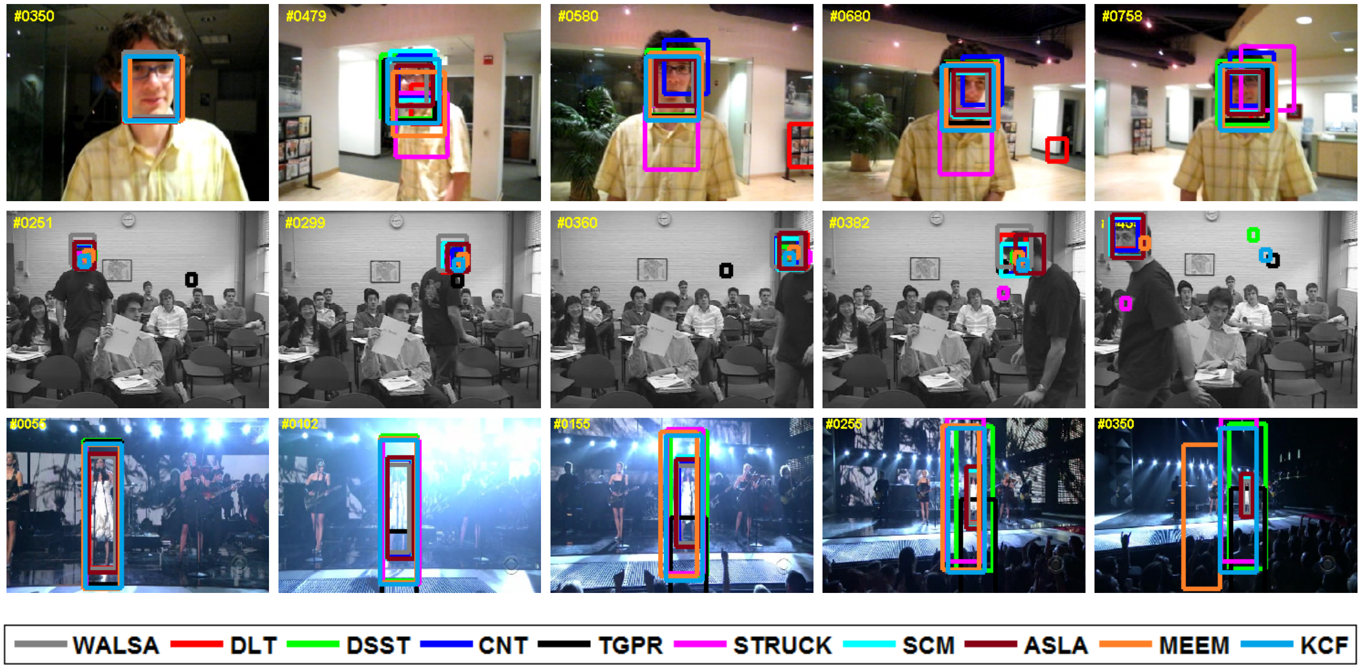


Fig. 7: Qualitative results of the 10 trackers over sequences *david*, *freeman3* and *singer1*, in which the targets undergo scale variations. Best viewed on color display.

[2] B. Ma, J. Shen, Y. Liu, H. Hu, L. Shao, and X. Li, "Visual tracking using strong classifier and structural local sparse descriptors," *IEEE Transactions on Multimedia*, vol. 17, no. 10, pp. 1818–1828, 2015. 1

[3] V. Kılıç, M. Barnard, W. Wang, and J. Kittler, "Audio assisted robust visual tracking with adaptive particle filtering," *IEEE Transactions on Multimedia*, vol. 17, no. 2, pp. 186–200, 2015. 1

[4] Y. Yuan, H. Yang, Y. Fang, and W. Lin, "Visual object tracking by structure complexity coefficients," *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1125–1136, 2015. 1

[5] C.-T. Chu, J.-N. Hwang, H.-I. Pai, and K.-M. Lan, "Tracking human under occlusion based on adaptive multiple kernels with projected gradients," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1602–1615, 2013. 1

[6] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, "Object tracking with multi-view support vector machines," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 265–278, 2015. 1

[7] K. Zhang and H. Song, "Real-time visual tracking via online weighted multiple instance learning," *Pattern Recognition*, vol. 46, no. 1, pp. 397–411, 2013. 1

[8] J. Tang, X. Shu, Z. Li, G.-J. Qi, and J. Wang, "Generalized deep transfer networks for knowledge propagation in heterogeneous domains," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 12, no. 4s, p. 68, 2016. 1

[9] J. Tang, L. Jin, Z. Li, and S. Gao, "Rgb-d object recognition via incorporating latent data structure and prior knowledge," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1899–1908, 2015. 1

[10] H. Song, "Active contours driven by regularised gradient flux flows for image segmentation," *Electronics Letters*, vol. 50, no. 14, pp. 992–994, 2014. 1

[11] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in *Proceedings of European Conference on Computer Vision*, pp. 188–203, 2014. 1, 5

[12] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," in *Proceedings of European Conference on Computer Vision*, pp. 188–203, 2014. 1, 5

[13] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3074–3082, 2015. 1, 2

[14] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Computer Vision-ECCV 2014 Workshops*, pp. 254–265, 2014. 1, 2

[15] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proceedings of European Conference on Computer Vision*, pp. 127–141, 2014. 1

[16] K. Zhang, L. Zhang, and M.-H. Yang, "Fast compressive tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 2002–2015, 2014. 1

[17] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1779–1792, 2016. 1, 4, 5

[18] H. Song, Y. Zheng, and K. Zhang, "Robust visual tracking via self-similarity learning," *Electronics Letters*, vol. 53, no. 1, pp. 20–22, 2016. 1

[19] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008. 1, 4, 5

[20] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proceedings of European Conference on Computer Vision*, pp. 234–247, 2008. 1

[21] H. Song, "Robust visual tracking via online informative feature selection," *Electronics Letters*, vol. 50, no. 25, pp. 1931–1933, 2014. 1

[22] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1838–1845, 2012. 1

[23] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1822–1829, 2012. 1, 2, 3, 4, 5

[24] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in Neural Information Processing Systems*, pp. 809–817, 2013. 1, 5

[25] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2259–2272, 2011. 1, 2, 4

[26] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1910–1917, 2012. 1

[27] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 821–829, 2018. 1

[28] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking

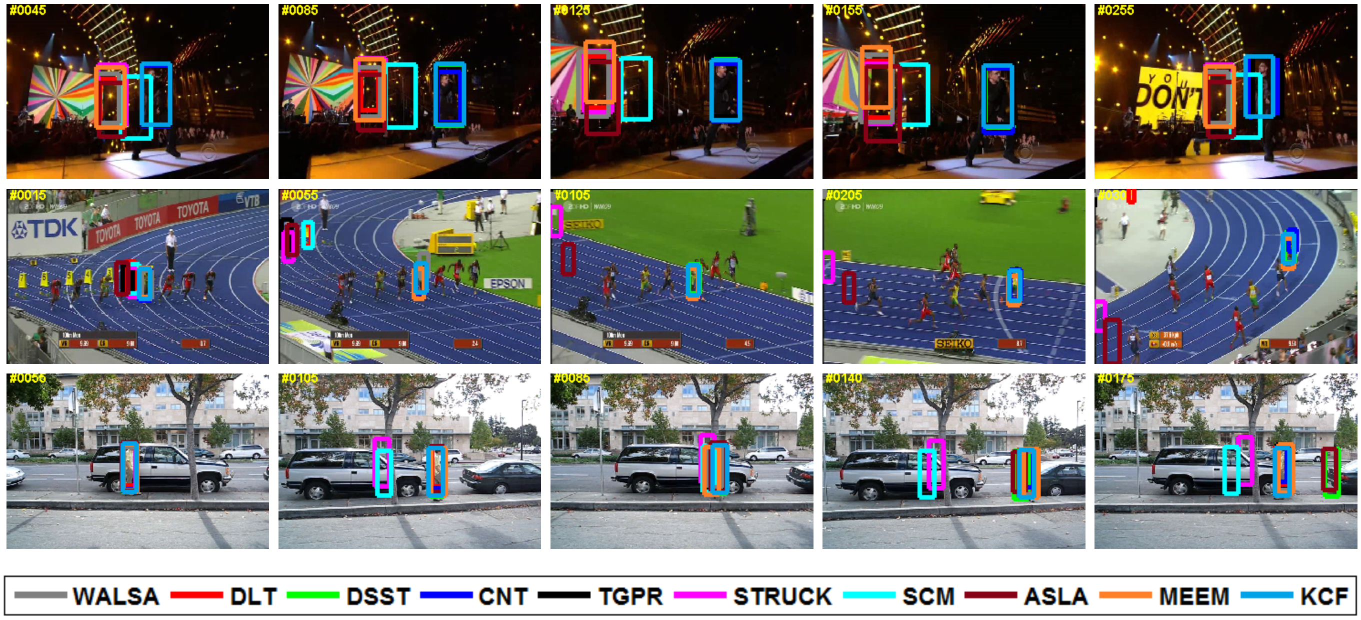


Fig. 8: Qualitative results of the 10 trackers over sequences *singer2*, *bolt*, and *david3* in which the targets undergo severe deformation. Best viewed on color display.

via multi-task sparse learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2042–2049, 2012. 1

[29] J. Kwon and K. M. Lee, "Tracking by sampling trackers," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1195–1202, 2011. 1

[30] S. Avidan, "Support vector tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064–1072, 2004. 1

[31] W. Chen, K. Zhang, and Q. Liu, "Robust visual tracking via patch based kernel correlation filters with adaptive multiple feature ensemble," *Neurocomputing*, vol. 214, pp. 607–617, 2016. 1

[32] J. Yang, K. Zhang, and Q. Liu, "Robust object tracking by online fisher discrimination boosting feature selection," *Computer Vision and Image Understanding*, vol. 153, pp. 100–108, 2016. 1

[33] H. Song, G. Wang, A. Cao, Q. Liu, and B. Huang, "Improving the spatial resolution of fy-3 microwave radiation imager via fusion with fy-3/mersi," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3055–3063, 2017. 1

[34] J. Tang, X. Shu, G.-J. Qi, Z. Li, M. Wang, S. Yan, and R. Jain, "Tri-clustered tensor completion for social-aware image tag refinement," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1662–1674, 2017. 1

[35] K. Zhang, L. Zhang, K.-M. Lam, and D. Zhang, "A level set approach to image segmentation with intensity inhomogeneity," *IEEE transactions on cybernetics*, vol. 46, no. 2, pp. 546–557, 2016. 1

[36] B. Zhang, A. Perina, Z. Li, V. Murino, J. Liu, and R. Ji, "Bounding multiple gaussians uncertainty with application to object tracking," *International journal of computer vision*, vol. 118, no. 3, pp. 364–379, 2016. 1

[37] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via structured multi-task sparse learning," *International journal of computer vision*, vol. 101, no. 2, pp. 367–383, 2013. 1

[38] T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem, "Robust visual tracking via consistent low-rank sparse learning," *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171–190, 2015. 1

[39] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, "Robust online appearance models for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1296–1311, 2003. 1

[40] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 798–805, 2006. 1

[41] K. Zhang, Q. Liu, H. Song, and X. Li, "A variational approach to simultaneous image segmentation and bias correction," *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1426–1437, 2015. 1

[42] K. Zhang, Q. Liu, J. Yang, and M.-H. Yang, "Visual tracking via boolean map representations," *Pattern Recognition*, vol. 81, pp. 147 – 160, 2018. 1

[43] J. Kwon and K. M. Lee, "Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive basin hopping monte carlo sampling," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1208–1215, 2009. 1

[44] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1631–1643, 2005. 1

[45] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via online boosting," in *Proceedings of British Machine Vision Conference*, pp. 47–56, 2006. 1

[46] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, 2011. 1

[47] Z. Kalal, J. Matas, and K. Mikolajczyk, "Pn learning: Bootstrapping binary classifiers by structural constraints," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 49–56, 2010. 1

[48] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 263–270, 2011. 1

[49] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015. 1, 2, 5

[50] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830–1837, 2012. 2

[51] B. Liu, J. Huang, L. Yang, and C. Kulikowsk, "Robust tracking using local sparse appearance model and k-selection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1313–1320, 2011. 2

[52] H. Li, C. Shen, and Q. Shi, "Real-time visual tracking using compressive sensing," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1305–1312, 2011. 2

[53] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2411–2418, 2013. 2, 5

[54] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering.," in *IJCAI*, pp. 3569–3575, 2015. 3

[55] M. J. Li, M. K. Ng, Y.-m. Cheung, and J. Z. Huang, "Agglomerative fuzzy k-means clustering algorithm with selection of number of clusters," *IEEE transactions on knowledge and data engineering*, vol. 20, no. 11, pp. 1519–1534, 2008. 3

[56] J. Xu, J. Han, K. Xiong, and F. Nie, "Robust and sparse fuzzy k-means clustering.," in *IJCAI*, pp. 2224–2230, 2016. 3

[57] L. Wen, Z. Cai, Z. Lei, D. Yi, and S. Z. Li, "Online spatio-temporal structural context learning for visual tracking," in *European Conference on Computer Vision*, pp. 716–729, Springer, 2012. 3, 4

[58] M. Yang, Y. Wu, and G. Hua, "Context-aware visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1195–1209, 2009. 4

[59] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015. 5

[60] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *British Machine Vision Conference*, 2014. 5

Zhetao Li is a professor in College of Information Engineering, Xiangtan University. He received the B.Eng. degree in Electrical Information Engineering from Xiangtan University in 2002, the M.Eng. degree in Pattern Recognition and Intelligent System from Beihang University in 2005, and the Ph.D. degree in Computer Application Technology from Hunan University in 2010. He is a member of IEEE and CCF. From Dec 2013 to Dec 2014, he was a post-doc in wireless network at Stony Brook University. From Dec 2014 to Dec 2015, he was an invited professor at Ajou University.His research interests include multimedia signal processing and machine learning.

Jie Zhang is a student in College of Information Engineering,XiangTan University.His research interests include Computer Vision and Pattern Recognition.

Kaihua Zhang is a Professor in the School of Information and Control, Nanjing University of Information Science & Technology, Nanjing, China. He received the B.S. degree in Technology and Science of Electronic Information from Ocean University of China (OUC) in 2006, the M.S. degree in Signal and Information Processing from the University of Science and Technology of China (USTC) in 2009 and Ph.D degree from the Department of Computing in the Hong Kong Polytechnic University in 2013. From Aug. 2009 to Aug. 2010, he worked as a Research Assistant in the Department of Computing, The Hong Kong Polytechnic University. His research interests include image segmentation, level sets, and visual tracking.

Zhiyong Li received the MSc degree in System Engineering from National University of Defense Technology, Changsha, China, in 1996 and PhD degree in Control Theory and Control Engineering from Hunan University, Changsha, China, in 2004. Since 2004, he joined the College of Computer Science and Electronic Engineering of Hunan University. Now, he is a Full Professor with Hunan University, member of IEEE and China Computer Federation (CCF). His research interests include visual object tracking, embedded computing system, dynamic multi-objective optimization evolutionary algorithm and tasks scheduling optimization in cloud computing.

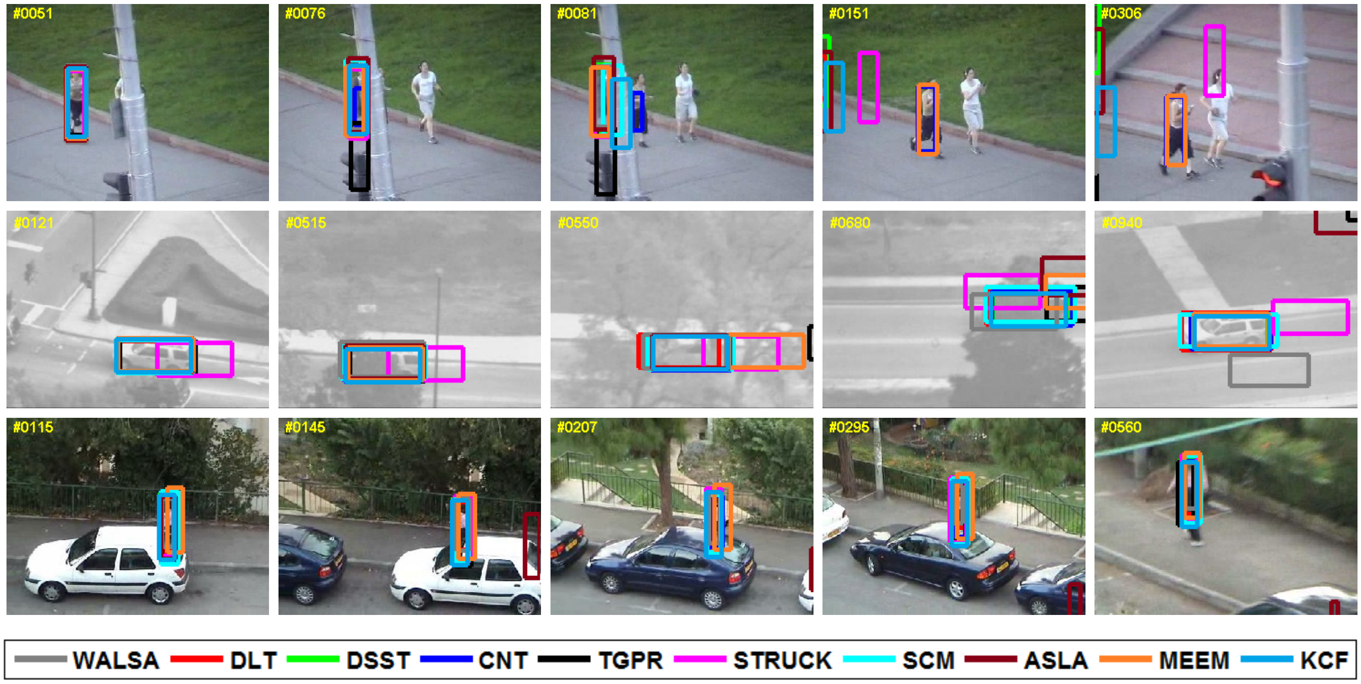


Fig. 9: Qualitative results of the 10 trackers over sequences *jogging-1*, *suv* and *woman*, in which the targets undergo heavy occlusion. Best viewed on color display.

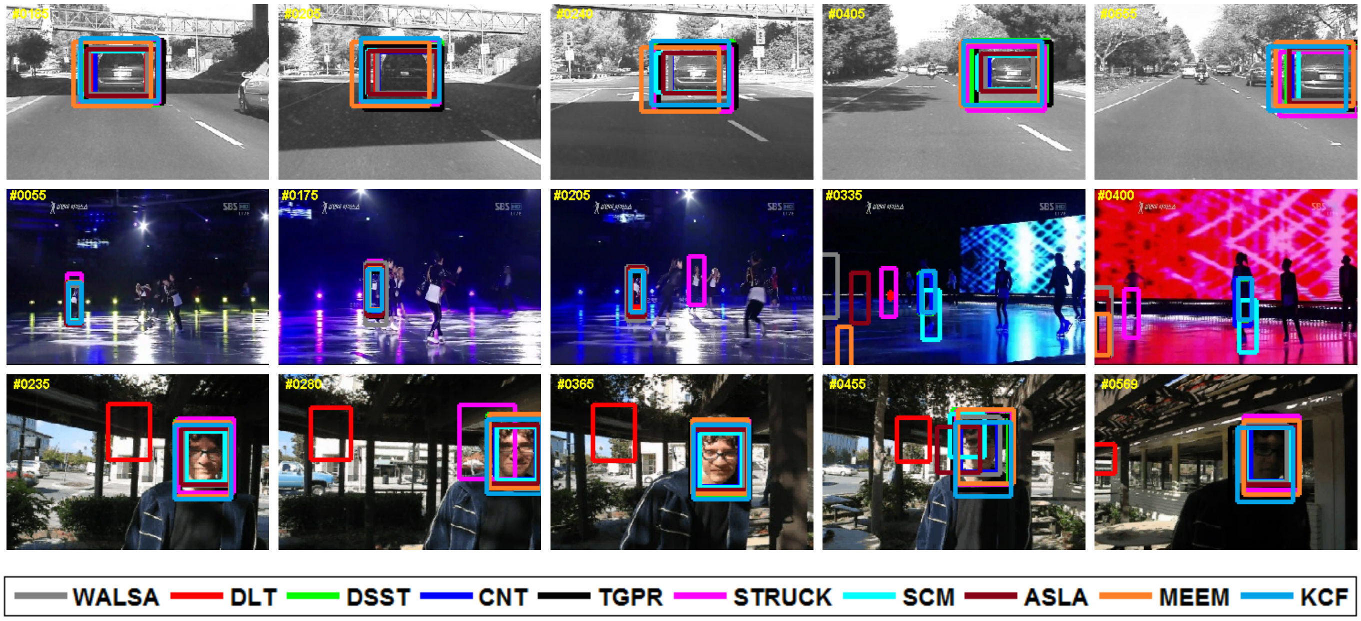


Fig. 10: Qualitative results of the 10 trackers over sequences *car4*, *skating1* and *trellis*, in which the targets undergo severe illumination changes. Best viewed on color display.

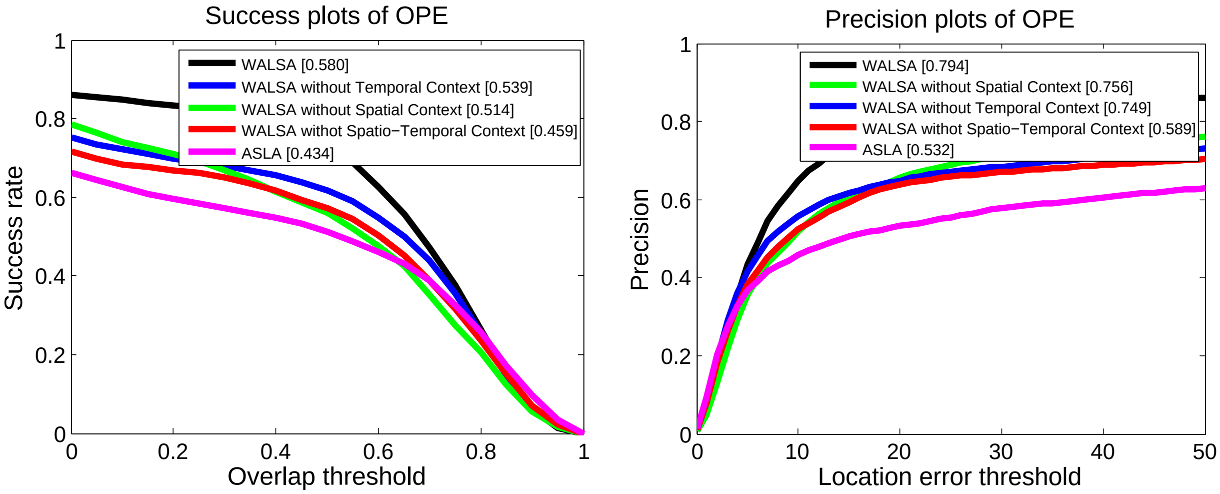


Fig. 11: Success plots and precision plots of OPE for WALSA with different components on OTB50. The ASLA is taken as a baseline. Best viewed on color display.