Enhanced Laplacian Group Sparse Learning with Lifespan Outlier Rejection for Visual Tracking

Behzad Bozorgtabar[1] and Roland Goecke[1,2]

[1]Vision & Sensing, HCC Lab, ESTeM
University of Canberra
[2]IHCC, RSCS, CECS
Australian National University
Behzad.Bozorgtabar@canberra.edu.au
roland.goecke@ieee.org

Abstract. Recently, sparse based learning methods have attracted much attention in robust visual tracking due to their effectiveness and promising tracking results. By representing the target object sparsely, utilising only a few adaptive dictionary templates, in this paper, we introduce a new particle filter based tracking method, in which we aim to capture the underlying structure among the particle samples using the proposed similarity graph in a Laplacian group sparse framework, such that the tracking results can be improved. Furthermore, in our tracker, particles contribute with different probabilities in the tracking result with respect to their relative positions in a given frame in regard to the current target object location. In addition, since the new target object can be well modelled by the most recent tracking results, we prefer to utilise the particle samples that are highly associated to the preceding tracking results. We demonstrate that the proposed formulation can be efficiently solved using the Accelerated Proximal method with just a small number of iterations. The proposed approach has been extensively evaluated on 12 challenging video sequences. Experimental results compared to the state-of-the-art methods demonstrate the merits of the proposed tracker.

1 Introduction

Object tracking is a well-studied problem in computer vision and has many practical applications. The problem and its difficulty depend on several factors, such as the amount of prior knowledge about the target object. Tracking of generic objects has remained challenging because an object can drastically change appearance when deforming (e.g. a pedestrian), rotating out of plane, being occluded, or when the illumination of the scene changes.

Recently, sparse representation has been strongly applied to visual tracking [1, 2]. In this case, the tracker represents each target candidate as a sparse linear combination of dictionary templates that can be dynamically updated to preserve an up-to-date target appearance model. However, sparse coding based trackers perform a computationally expensive l_1 minimisation at each frame.

The drawback of these methods is that they ignore the underlying structure between particles and learn sparse representations of particles separately. Ignoring the relationships among particle representations tend to make the tracker more prone to drifting away from the target, especially in cases of significant appearance changes of the tracking target.

In this paper, we propose a computationally efficient Laplacian group sparse learning approach for visual tracking in a particle filter framework. Here, we consider each particle sample as a task and explore task correlation between particles. Besides, we further extend our designed objective function with the Laplacian norm to recognise the overall structure among particles with a defined similarity graph. Unlike previous methods, we also consider the consistency between tracking results in a short period of time and since the new tracking result is more likely to be similar with the most recent ones, we construct another set of dictionary items, but this time for the current candidate samples, select the most correlated ones and ignore the rest.

Finally, a new likelihood model is proposed based on two factors: **(1)** the relative location of the particles w.r.t. the current target object and **(2)** the similarity between target candidates and the dictionary templates. Consequently, once the current tracking result causes a large variance and is not occluded, we replace it with the dictionary template has less similarity with the tracking result.

2 Related Work

In general, object tracking methods can be categorised as either generative or discriminative.

2.1 Generative Trackers

Generative methods represent the target object with models that have minimum reconstruction errors, and track targets by searching for the region most similar to the models in an image frame. Examples of generative methods are eigentracker [3], mean shift tracker [4], context-aware tracker [5], fragment-based tracker (Frag) [6], incremental tracker (IVT) [7], and VTD tracker [8]. Most recent generative methods learn and maintain static or online appearance models. Black et al. [3] learn a subspace model offline to represent target objects at predefined views and build on the optical flow framework for tracking.

2.2 Discriminative Trackers

Discriminative models, which are also called tracking-by-detection methods, consider tracking as a binary classification task to separate the object from its surrounding background. The adaptive tracking-by-detection methods first train a classifier in an online manner using samples extracted from the current frame. In the next frame, a sliding window is then used to extract samples around the previous object location, before the previously trained classifier is applied to these samples. The location of the sample with the maximum classifier score is the new object location at the current frame. Examples of discriminative methods are on-line boosting (OAB) [9], ensemble tracking [10], co-training tracking [11], adaptive metric differential tracking [12] and online multiple instance learning tracking [13].

2.3 Sparse Representation for Object Tracking

The recent development of sparse representations [1, 2] has attracted considerable interest in object tracking due to its robustness to occlusion and image noise. In [1], a target candidate is represented as a sparse linear combination of object templates and trivial templates. For each particle, a sparse representation is computed by solving a constrained l_1 minimisation problem with non-negativity constraints, thus, solving the inverse intensity pattern problem during tracking. Although this method yields good tracking performance, it comes at the computational expense of multiple l_1 minimisation problems that are independently solved. In Mei et al. [2], an efficient l_1 tracker with minimum error bound and occlusion detection is proposed. Zhang et al. [14] investigate convex mixed norm $l_{p,q}$ (*i.e.* $p \geq 1, q \geq 1$) to enforce joint sparsity for the particles. In [15], a particle filter based tracking formulated as a structured multi-task sparse learning problem, where particle representations, regularised by a sparsity-inducing mixed norm and a local graph term.

3 System Overview

3.1 Bayesian Inference Framework

In this paper, visual tracking is formulated within the Bayesian inference framework, in which the goal is to determine the *a posteriori* probability of the target state. In this paper, we utilise the particle filter as an effective realisation of Bayesian filtering, whereas the idea is to approximate the posterior distribution $p\left(s_t \left| z_{1:t}\right.\right)$ by a set of weighted particles $\left\{s_t^{(i)}, \pi_t^{(i)}\right\}_{i=1}^{N}$ where $z_{1:t}$ denotes the set of observations up to and including the time step t and each particle represents a possible state s_t and a weight π_t associated with it, which specifies its corresponding state's confidence. Considering Bayesian estimation scheme, the filtering distribution can be recursively updated as:

$$p\left(s_t \left| z_{1:t-1}\right.\right) = \int p\left(s_t \left| s_{t-1}\right.\right) p\left(s_{t-1} \left| z_{1:t-1}\right.\right) ds_{t-1} \tag{1}$$

$$p\left(s_t \left| z_{1:t}\right.\right) \propto p\left(z_t \left| s_t\right.\right) p\left(s_t \left| z_{1:t-1}\right.\right) \tag{2}$$

First, new particles are generated by sampling from a known proposal function $q\left(s_t \left| s_{0:t-1}^{(i)}\right., z_{1:t}\right)$ where the simplest choice for the proposal function is the state evolution model $p(s_t | s_{t-1})$ itself for sampling. Further, the optimal state is obtained by the maximum a posteriori (MAP) estimation over a set of N samples. In our algorithm, we model the motion of a target object between two consecutive frames with an affine transformation. Let s_t be the six-dimensional parameter vector of an affine transformation. The transformation of each parameter is modelled independently by a scalar Gaussian distribution. Then, the dynamic model $p(s_t | s_{t-1})$ can be represented by a Gaussian distribution. The likelihood (observation) model $p(z_t | s_t)$ reflects the similarity measure for the tracking target. In this paper, the weights of the particles are specified by the proposed spatial score weighted by the difference in contribution of object templates and the background templates, in which a sample with a larger difference score indicates that it is more likely to be correlated with target object rather than background. The most likely sample is considered as the tracking result for that video frame.

3.2 Contribution

Inspired by the mentioned related works, we seek a particle filter based tracker using sparse representation scheme, which not only reduces the computational expense caused by regressing individual particles with respect to the defined dictionary, but also models the common structure among particle samples. Zhang et al. [15] extend the MTT framework to take into account pairwise structural correlations between particles. However, our tracker is superior and more stable. The key idea is constructing a similarity graph to better model structural information of the sampled particles. To represent the graph reqularizer, the tracker in [15] only uses pairwise distance between each pair of particles by considering their spatial locations, which ignores the global structure of the whole particles and is prone to the outliers. In [16], it is extensively shown that each data point in a union of subspaces can be efficiently reconstructed by a combination of other points in the dataset. However, the spanned subspaces are usually dependent, which causes the wrong choice of inter subspaces. Based on this assumption, we consider both the mutual local and global structure of the particles. Here, we summarised our main contributions:

1. In this paper, inspired by the similar successful works in image classification e.g [17], we formulate object tracking by proposing an enhanced Laplacian group sparse coding based scheme where the similarity among the particles specified by a graph structure, which makes the sparse codes of those particles placed close together be similar to each other (e.g. spatial smoothness). Furthermore, we investigate an enhanced similarity graph which not only considers pairwise similarity between particles, but also encode linearly representation of each particle with respect to others in a topological space. The proposed objective function for learning the sparse representation of candidate samples is rendered as a non-smooth convex (unconstrained) optimisation problem in which we implement the accelerated proximal method (APM) for solving this optimisation problem and develop efficient algorithms for computing the related proximal mapping.

2. The target object position ambiguity problem often occurs in visual tracking, which adversely influences tracking performance. The proposed tracker incorporates the particle sample importance into the observation model, which makes it able to select the most effective particles, resulting in a more stable tracker.
3. Finally, for the sake of efficient computational complexity and having a faster tracker, we introduce outlier rejection for the particle samples in the current frame where we take the advantage of this fact that the current tracking result can be well represented by the latest tracking results and pick those, which are more similar to the previous ones and more likely to be the new tracking result.

In our particle filter based tracking method, particles are randomly sampled around the current state of the tracked object according to a zero mean Gaussian distribution. In the t^{th} frame, we consider N particle samples, whose observations (Gray scale values) are denoted in matrix form as: $Y = [y_1, \ldots, y_N] \in R^{m \times N}$ [1]. We construct our dictionary $D_t = [d_1, \ldots, d_K] \in R^{m \times K}$, in which the tracked object can be represented under a variety of appearance changes by its templates $\{d_i\}_{i=1}^{K}$ (d_i is the i^{th} dictionary item).

3.3 Joint Sparse Model

Since in particle filter based visual tracking, particle are densely sampled around the current target state, there are often underlying correlation structure between the particles. To explore these hidden structures, [14] employed multi task sparse learning to impose joint sparsity between the particles (tasks) yields a more accurate representation for the ensemble of particles where the sparse representation matrix $X = [x_1, \ldots, x_N] \in R^{K \times N}$ can be obtained by as follows:

$$\min_{X} \|Y - D_t X\|_F^2 + \lambda \|X\|_{p,q} \tag{3}$$

Where $\|X\|_{p,q} = \left(\sum_{i=1}^{K} \|X_i\|_p^q\right)^{\frac{1}{q}}$, $\|X_i\|_p$ is the l_p norm of X_i, (i^{th} row of matrix X).

3.4 Laplacian Group Sparse Model

As has been discusse discussed, in order to alleviate computationally expense caused by l_1 minimisation of each particle separately in particle filter based tracking, we seek the common structure among tasks (particle samples) in any given frame. However, considering a global structure for particles is not strong assumption. In fact, in practical application, the particles may exhibit a more sophisticated structure where the sparse representation of closely particles is more likely to be similar rather than those from different spatial locations. In

[1] $m = 1024 - dim$ Gray scale based features.

this paper, we augment multi task learning framework with a graph structure to consider mutual relation between particles. In this way, is each frame we construct the similarity graph where the particles are represented by the nodes and the edges between particles specify their correlation (feature similarity). Motivated by the success of recent work in image classification[17], we have the following representation:

$$\|Y - D_t X\|_2^2 + \frac{\lambda_1}{2} Tr\left(X\hat{L}X^T\right) + \lambda_2 \|X\|_{p,1} \tag{4}$$

where $\|X\|_{p,1} = \sum_{i=1}^{K} \|X\|_p$, λ_1 and λ_2 are the regularisation parameters. $\hat{L}$ known as the Laplacian matrix, is symmetric and positive definiteness and acts as a key factor in our proposed tracker that models the similarity graph for the particles.

3.5 Solving the Optimisation Problem

The formulated problem in Eq. 4 is a convex optimisation problem with a non-smooth objective function due to the non-negativity constraint assumption for the particles representation matrix X. In this paper, we seek to solve this optimisation problem using the accelerated proximal method (APM) [18] due to its ability of optimal convergence compared to other first-order techniques. APM iterates between two sequences of variables: **(1)** an attainable solution (updating the current representation matrix) $\left\{\hat{X}_k\right\}$ and **(2)** an aggregation matrix sequence $\left\{\nabla^k\right\}$.

Proximal mapping: At each iteration, the representation matrix X_k can be updated by the generaliSed proximal mapping as the following problem:

$$\min_X \frac{1}{2} \|X - H\|_2^2 + \widetilde{\lambda} \|X\|_{p,1} \tag{5}$$

where $\widetilde{\lambda} = \frac{\lambda_2}{\gamma_k}$, $H = \hat{X}_k - \frac{1}{\gamma_k}\nabla^k$ and γ_k denotes the step size.

$$\nabla^k = \hat{X}_k - \frac{1}{\gamma_k}\left(D_t^T\left(Y - D_t\hat{X}_k\right) + \lambda_1 \hat{X}_k \hat{L}\right) \tag{6}$$

Aggregation sequence: At the k^{th} iteration of APM, the aggregation matrix is updated by linear combination of X_k and X_{k-1} from previous iterations: [2]

$$\hat{X}_{k+1} = X_{k+1} + \frac{\mu_{k+1}(1-\mu_k)}{\mu_k}(X_{k+1} - X_k) \tag{7}$$

[2] μ_k is conventionally set to $\frac{2}{k+1}$.

3.6 Enhanced Similarity Graph for the Particles

Regarding the graph structure for the particle samples in the current frame, the most existing similarity methods only consider pairwise distances of the data points. Since the pairwise distance only rely on the two connected samples (nodes) and ignores the global structure of the whole sample points, it is fragile to outliers. To address this problem, we propose a similarity graph, which takes advantage of a linear representation of each particle over a set of other particle samples. The proposed similarity graph not only could encode each particle feature over the other particles with less residual error, but also enforces the local representation of each particle sample using the following objective function:

$$\min_{c_i} \sum_{i=1}^{N} \rho \left\| S_i c_i \right\| + (1-\rho) \left\| y_i - A_i c_i \right\| \quad s.t. \mathbf{1}^T c_i = 1 \tag{8}$$

where we encode each particle sample by building the dictionary A_i composed of the set of remaining particles at the given frame, $A_i = [y_1, y_2, \cdots, y_{i-1}, 0, y_{i+1}, \cdots, y_N]$, S_i is a diagonal matrix whose j^{th} diagonal element is the pairwise distance from y_i to the j^{th} particle in A_i,$(j = 1, 2, \cdots, N)$, c_i is a similarity over the corresponding dictionary, $\mathbf{1} \in \Re^N$ (N is the number of candidates) denotes the column vector whose entries are all ones and and ρ is the factor controls the balance between linear representation measure and pairwise distance scheme. The similarity representation of the proposed graph can be solved as:

$$c_i = \frac{R_i^{-1}\mathbf{1}}{\mathbf{1}^T R_i^{-1}\mathbf{1}} \tag{9}$$

where $R_i = (1-\rho)\left(y_i \mathbf{1}^T - A_i\right)^T \left(y_i \mathbf{1}^T - A_i\right) + \rho S_i^T S_i$. Furthermore, in order to obtain a more discriminative correlation matrix, we keep the k strongest connections for each particle using k-nearest neighbour searching on c_i and set all other elements to zero. At the end, we use the obtained similarity matrix as our Laplacian matrix $\hat{L}$ in Eq. 4.[3]

4 Lifespan Outlier Rejection

Since the appearance of a target object is expected to be temporally correlated and does not change dramatically over a short period of time, the target object can be represented well by the more recent tracking results w.r.t. the current frame. Therefore, in each frame, we aim to select the particle samples that are highly correlated to the previous tracking results and remove the unrelated ones. In order to achieve this goal, we construct another set of dictionary templates

[3] We denote $\widetilde{c}_i$ as a discriminative feature, then we build a similarity graph by considering each point as a vertex and assigning the connection weight between the node i and j as $|\widetilde{c}_{ij}|$.

$T = \{t_i\}_{i=1}^{n}$, in which each column represents the tracking result of the latest $n-1$ frames[4] plus the template in the first frame, which is considered as our proposal. Given the new template set, the sparse coefficient w_i for each template t^i can be computed by:

$$\min_{w_i} \left\| t^i - Y w_i \right\|_2^2 + \lambda_w \left\| w_i \right\|_1 s.t. \forall i, w_i \geq 0, i = 1, \cdots, n \tag{10}$$

where λ_w is the regularisation factor. Using l_1 minimisation of the LARS algorithm [19], we are capable of choosing the highly associated particle samples for the current frame with reduced computational complexity. The implementation also has an option to add positivity constraints on the solutions w_i, which give us a precise sparse representation. This representation reveals a small or even zero value for the j^{th} candidate if it holds little similarity with the i^{th} template. Therefore, we build a mask for each particle sample, in which if the sum of its sparse coefficients is zero, we set its weight ω_j to zero and ignore this particle as follows:

$$\omega_j = \begin{cases} 0 & if \sum_i^n w_{ij} = 0 \\ 1 & \text{Otherwise} \end{cases} \tag{11}$$

In this way, we build a faster tracker with the least number of particles as possible (see Fig. 1). Consequently, the selected particle samples will adaptively change according to different scenes.

5 Likelihood Model Using the Spatial Weight

We present a spatial confidence score for the particles, which can naturally integrate the particle sample importance into the tracking result. Here, we assume the tracking location at the current frame is the location of the most correct particle sample to make each particle contribute differently to the target presence probability: The closer the location is to the current tracked position, the larger probability it has and the farther it is from the tracking location, the less it contributes to the object presence probability (see Fig.2).

Up to now, the proposed tracker is entirely generative. However, in order to handle the drifting problem caused by rapid changes in appearance of the tracking object, we build up target templates with background templates randomly sampled in the first frame, which consequently should be updated in successive frames. The spatial score for the i^{th} particle can be modelled as:

$$\psi_i = \frac{1}{C} e^{-|l_i - l^*|} \tag{12}$$

where C is the normalisation constant and $l(\cdot) \in \Re^2$ is the location function. This spatial weight is a monotone decreasing function w.r.t. the Euclidean distance between the locations of the i^{th} particle sample and the target location l^*.

[4] We consider every $n = 5$ frames.

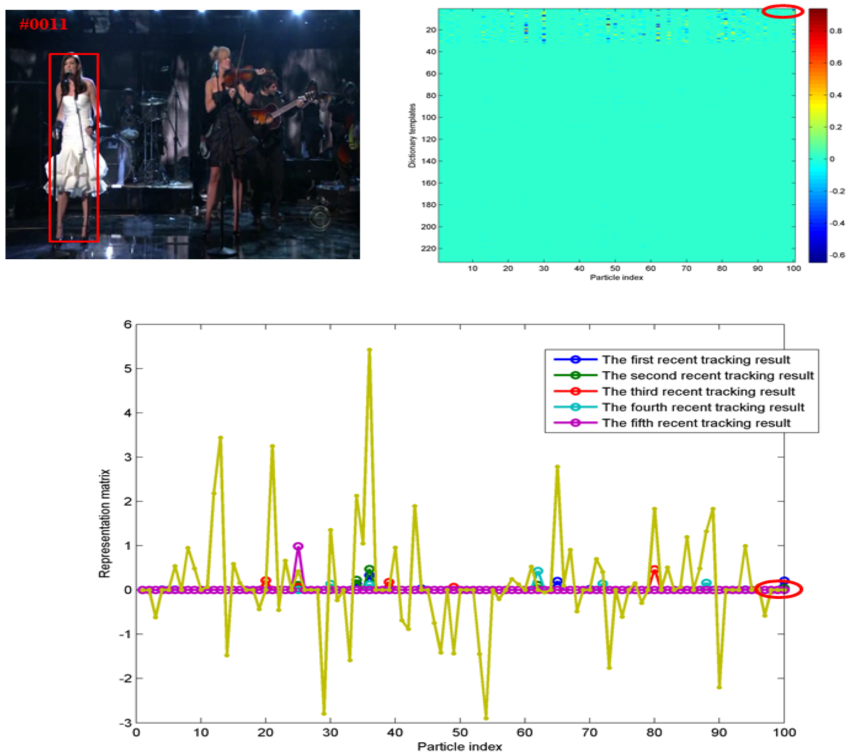

Fig. 1. Illustrating the effect of outlier rejection for the *singer 1* sequence. At the top right, the sparse representation of particles with respect to the target templates is shown, while brighter columns indicate the presence of outliers. At the bottom, sparse representations of the most recent tracking results w.r.t. the current particles are shown. As highlighted, those particles whose sum of their sparse representations is zero are removed from sampling.

Finally, in order to obtain the overall score for the tracking result, we first divide the sparse representation matrix X into two subsets, X_{pos} and X_{neg}, each representing the candidate's similarity to the object (positive) templates $D_t^{(O)}$ and background (negative) templates $D_t^{(B)}$,respectively. Further, the weights of the particles are specified by the difference in contribution of these two parts and the tracking result z_t at time instance t is the particle y_i such that:

$$i = \arg\max y_{i=1,\ldots,N} \quad \psi_i \odot \left(\|X_{pos}\|_1 - \|X_{neg}\|_1\right) \tag{13}$$

where $\odot$ is an element-wise product[5]. This likelihood function not only encourages the tracking result to be represented well by the object and *not* the back-

[5] This is the element-wise product of two $(1 \times N)$ matrices. N is the number of sampled particles.

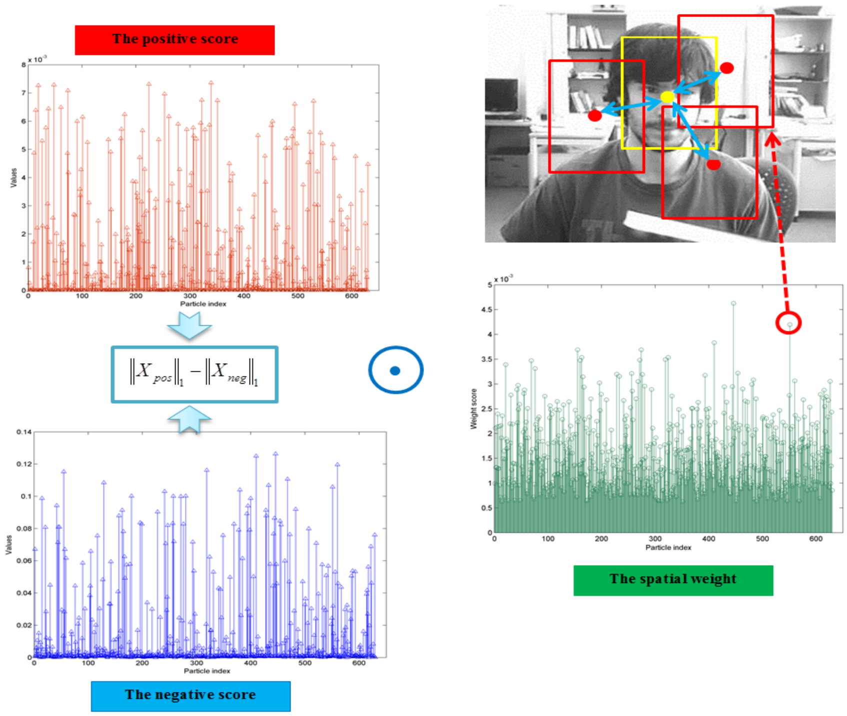

Fig. 2. This figure illustrates the principle of the proposed likelihood model. On the left side, the discriminative scores of candidates are shown. On the right side, the relative distance between particles and target object is illustrated. The yellow rectangle is the tracking result. The solid circles are the central locations of each particle and red rectangles are the particle samples. The corresponding spatial weight of the sample particle is highlighted.

ground templates, it also gives more weight to the particle samples near the current tracking location (Fig. 2).

6 Dictionary Update

Updating target object templates for handling appearance change during tracking is a vital part of any visual tracking method. Neither fixed object templates nor frequently updated target templates could help the accurate representation of target appearance during tracking. Since target appearance only remains unchanged in just a short period of time, a stable appearance model is not reliable for long period tracking. On the other hand, if samples are updated frequently, the model will degrade. The initial dictionary comprising positive templates n_p is obtained by drawing sample images around the target location (e.g. within a radius of a few pixels) and downsampling the selected images to a normalised size (32×32 pixels in our experiments). Each downsampled image is stacked together to form the set of positive templates. Similarly, the set of n_n negative templates is composed of images further away from the labelled location (e.g. within an annular region some pixels away from the target object).

First, for the background templates $D_t^{(B)}$, since the surrounding image region of two consecutive frames are similar, we sample the negative templates from image regions away (e.g. more than 7 pixels) only from the current tracking result. Moreover, we take advantage of using our proposed target transition model to ignore those candidates as tracking result since they are involved in representing negative templates.

On the other hand, for the object template set $D_t^{(O)}$, we allocate a similarity measure μ_i that demonstrates how representative the template in tracking result is. The more a template is used to represent tracking results, the higher its weight is. Similar to [20], in each frame, we measure the similarity between the current tracking result and the object templates and if the particles are not sufficiently represented (up to a predefined threshold $\Omega = 0.4$) by the dictionary, we use the tracking result to replace the corresponding template with the new tracking result. In other words, the tracking result is added to the template set if none of the template is comparable to the tracking result. The weight of this new template is set to the median of the current normalised weight vector.

7 Experiments and Results

In this paper, we compare the performance of the proposed tracker with several state-of-the-art trackers. We use the default parameters for these trackers as they reported. These trackers can be categorized to different groups. Discriminative heuristic trackers include the STRUCK [21], the compressive tracking based CT [22] and the multiple instance learning-based tracker MIL [23]. On the opposite side, generative trackers such as the incremental subspace based IVT [7] and two channels blurring approaches DFT [24] are considered. We also utilise part-based trackers akin to the TLD [25], which estimate the new target object by combining the local motion estimates with discriminative learning of patches and Frag-Track [6], in which the target object is represented by multiple image fragments or patches. In addition, sparse representations based trackers like ASLA [26], SCM [27], MTT [14] and L1APG [28] are used. Theses trackers range from local sparse representations (ASLA tracker) to holistic sparse templates (MTT and L1APG) and both local-holistic representation method (SCM). Finally, we implement the VTD method as our last benchmark, which adapts mixture models based on sparse principal component analysis.

7.1 Parameters Setting

The parameters of the proposed tracking algorithm are fixed in all experiments. The numbers of positive templates n_p and negative templates n_n are 50 and 200, respectively. All weight parameters of Eq. 4 are set to 0.5 and the regularisation parameter λ_w in Eq. 10 is fixed to be 0.01. In Eq. (6), we set $\widetilde{\lambda}$ (by cross-validation) to 0.005 and γ_k to 1/0.01, respectively. ρ in Eq. (8) is set to 0.5. The maximum iteration of the objective function in Eq. 4 is set to 10, and 100 particles are chosen as candidate samples in each frame. An observed target image patch is partitioned into non-overlapping local fragments (image patches) of size 8×8 pixels, each of which is independently represented in gray scale values, vectorised and normalised to be a vector with unit l_2 norm. Then, we concatenate these local feature vectors so that the global structural information is maintained. The candidates and templates in this work are all represented with this locally normalized features to handle partial occlusion and to moderate appearance variation.

7.2 The Test Sequences

We use 12 challenging video sequences widely used in the literatyre and publicly available from the online object tracking benchmark[6]. Each video sequence was labelled with different attributes including five visual attributes that reflect a specific challenge in appearance humiliation: **(i)** abrupt motion, **(ii)** illumination change, **(iii)** occlusion, **(iv)** scale change and **(v)** camera motion. Their ground truths are provided. Figure 3 shows some tracking results for different video sequences.

7.3 Performance Measure

Numerous performance metrics have been proposed for visual tracking evaluation during recent years. For the purpose of measuring the performance of the proposed tracker, two criteria, the centre location error as well as the overlap ratio, are implemented here. It should be noted that a smaller average error or a bigger overlap rate means a more accurate result. The tracker's overlap rate in each frame defined as the area $\frac{area(BB_T \cap BB_G)}{area(BB_T \cup BB_G)}$, where BB_G and BB_T denote the bounding box obtained by the ground truth and a tracker, respectively. An important advantage of the overlap measure is that it accounts for both position and size of the predicted and proposal bounding boxes simultaneously and does not lead to arbitrary large errors at tracking failures. As is shown in Tables 1 and 2, except for two sequences, the results of our tracker outperform the other trackers. Due to space limitations, we do not mention the detailed results including the average error plots and average overlap ratio plots here but do so in the supplementary material.

8 Conclusion

In this paper, we have presented an enhanced Laplacian group sparse learning method for particle filter based visual tracking. By imposing the Laplacian group

[6] `https://sites.google.com/site/trackerbenchmark/benchmarks/v10`

Table 1. Average centre location errors (in pixels). The best three results are shown in Red, Blue, and Green fonts.

Video Clip	Frag	IVT	MIL	APG	VTD	MTT	SCM	CT	TLD	ASLA	Struck	DFT	Our
Singer1	22.0	8.5	15.2	3.2	4.1	41.2	3.8	25.1	11.6	5.3	12.6	10.4	2.8
Girl	18.0	48.6	32.2	62.4	21.4	23.9	9.7	21.0	20.3	12.5	10.0	21.5	9.3
Car11	63.8	2.1	43.5	1.7	27.1	1.8	4.1	6.0	25.1	2.0	1.9	2.2	1.5
Face	48.8	69.7	134.6	57.7	140.9	127.2	125.1	144.2	67.5	95.1	25.0	26.8	15.3
David	76.7	3.6	16.1	14.3	13.6	124.2	5.1	15.3	16.3	6.0	3.1	10.2	3.4
Dudek	61.5	8.8	20.3	70.6	66.0	53.8	9.2	23.1	10.5	10.6	11.5	9.5	8.6
Woman	113.6	167.4	122.3	118.5	136.6	127.2	4.3	109.6	110.4	3.2	10.1	15.3	3.1
Bolt	240.1	170.6	163.9	225.5	22.3	106.0	73.2	115.5	34.5	56.4	98.5	102.3	16.5
Jumping	58.6	36.7	10.2	9.1	63.2	19.3	3.9	9.0	8.0	39.2	42.0	39.5	4.6
Mountain	141.6	33.2	128.3	130.2	7.5	11.3	5.9	86.7	96.5	5.1	10.5	122.4	3.3
Sylvester	98.9	70.8	31.1	112.5	49.4	14.6	9.0	21.3	17.5	9.1	20.4	34.5	8.5
Tiger1	39.5	158.7	14.2	21.5	28.9	30.9	10.5	20.0	13.9	11.4	12.2	10.0	8.9

Table 2. Average overlap rate (in pixels). The best three results are shown in Red, Blue, and Green fonts.

Video Clip	Frag	IVT	MIL	APG	VTD	MTT	SCM	CT	TLD	ASLA	Struck	DFT	Our
Singer1	0.34	0.66	0.33	0.83	0.79	0.32	0.85	0.29	0.65	0.78	0.59	0.72	0.87
Girl	0.69	0.43	0.51	0.33	0.52	0.63	0.69	0.78	0.57	0.72	0.94	0.56	0.96
Car11	0.09	0.81	0.17	0.83	0.43	0.58	0.79	0.71	0.38	0.82	0.86	0.63	0.91
Face	0.39	0.44	0.15	0.35	0.24	0.26	0.36	0.13	0.46	0.21	0.78	0.75	0.86
David	0.19	0.71	0.45	0.57	0.53	0.28	0.69	0.25	0.44	0.63	0.79	0.61	0.81
Dudek	0.46	0.81	0.64	0.61	0.46	0.36	0.76	0.51	0.71	0.73	0.75	0.68	0.83
Woman	0.20	0.18	0.16	0.06	0.15	0.17	0.77	0.16	0.07	0.81	0.86	0.74	0.85
Bolt	0.07	0.13	0.16	0.10	0.82	0.19	0.29	0.12	0.77	0.31	0.15	0.11	0.92
Jumping	0.13	0.29	0.54	0.57	0.09	0.31	0.72	0.96	0.98	0.25	0.18	0.20	0.98
Mountain	0.06	0.66	0.14	0.11	0.89	0.81	0.91	0.11	0.25	0.92	0.87	0.10	0.96
Sylvester	0.06	0.51	0.54	0.28	0.44	0.52	0.89	0.65	0.68	0.88	0.70	0.62	0.91
Tiger1	0.19	0.71	0.39	0.15	0.73	0.75	0.82	0.53	0.65	0.80	0.73	0.89	0.95

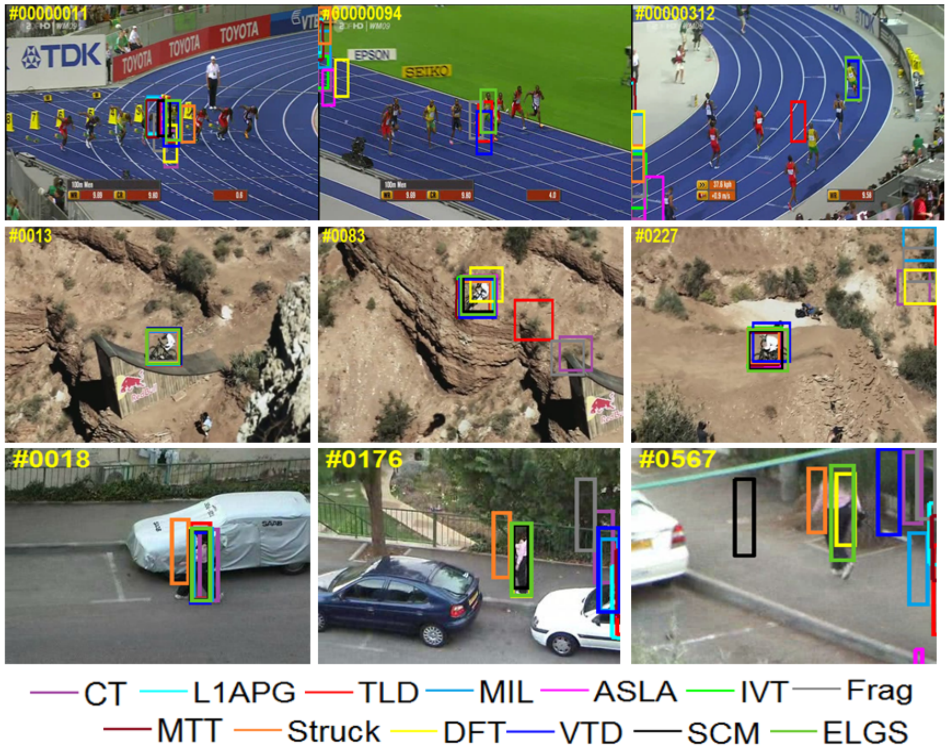

Fig. 3. Sample tracking results by our tracker (Enhanced Laplacian Group Sparse (ELGS)) compared with benchmark results in challenging sequences with rotation and non-rigid deformation (*Bolt sequence*), background clutter (*MountainBike sequence*) and heavy occlusion (*Woman sequence*).

sparse penalty term in our objective function, we are able to not only exploit the underlying relationship shared by different particles, but also to capture their structure ignored by previous works. Furthermore, we propose an enhanced similarity graph for the particle samples robust to outliers. In addition, since the target object can be modelled well by the more recent tracking results, in each frame, we remove the particle samples that are not correlated to the previous tracking results. Finally, a new likelihood function using the discriminative weighted particles is proposed where the particle importance is considered. In comparison with 12 state-of-the-art trackers, our tracker shows superior performance in both accuracy measures used.

References

1. Mei, X., Ling, H.: Robust visual tracking and vehicle classification via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011) 2259–2272
2. Mei, X., Ling, H., Wu, Y., Blasch, E., Bai, L.: Minimum error bounded efficient ? 1 tracker with occlusion detection. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2011) 1257–1264
3. Black, M.J., Jepson, A.D.: Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In: Computer VisionECCV'96. Springer (1996) 329–342
4. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 564–577
5. Yang, M., Wu, Y., Hua, G.: Context-aware visual tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009) 1195–1209
6. Adam, A., Rivlin, E.: Robust Fragments-based Tracking Using the Integral Histogram. In: Computer Vision and Pattern Recognition (CVPR). (2006) 798–805
7. Ross, D., Lim, J., Lin, R.S., Yang, M.H.: Incremental Learning for Robust Visual Tracking. International Journal of Computer Vision **77** (2008) 125–141
8. Kwon, J., Lee, K.M.: Visual tracking decomposition. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2010) 1269–1276
9. Grabner, H., Grabner, M., Bischof, H.: Real-Time Tracking via Online Boosting. In: British Machine Vision Conference (BMVC). (2006) 47–56
10. Avidan, S.: Ensemble tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007) 261–271
11. Liu, R., Cheng, J., Lu, H.: A robust boosting tracker with minimum error bound in a co-training framework. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE (2009) 1459–1466
12. Jiang, N., Liu, W., Wu, Y.: Adaptive and discriminative metric differential tracking. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2011) 1161–1168
13. Babenko, B., Yang, M., Belongie, S.: Visual Tracking with Online Multiple Instance Learning. In: Computer Vision and Pattern Recognition (CVPR). (2009)
14. Zhang, T., Ghanem, B., Liu, S., Ahuja, N.: Robust visual tracking via multi-task sparse learning. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2012) 2042–2049
15. Zhang, T., Ghanem, B., Liu, S., Ahuja, N.: Robust visual tracking via structured multi-task sparse learning. International journal of computer vision **101** (2013) 367–383
16. Elhamifar, E., Vidal, R.: Sparse subspace clustering: Algorithm, theory, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 2765–2781
17. Gao, S., Tsang, I.H., Chia, L.T.: Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 92–104
18. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences **2** (2009) 183–202
19. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al.: Least angle regression. The Annals of statistics **32** (2004) 407–499

20. Zhang, T., Ghanem, B., Liu, S., Ahuja, N.: Low-rank sparse learning for robust visual tracking. In: Computer Vision–ECCV 2012. Springer (2012) 470–484
21. Hare, S., Saffari, A., Torr, P.H.: Struck: Structured output tracking with kernels. In: 2011 IEEE International Conference on Computer Vision (ICCV), IEEE (2011) 263–270
22. Zhang, K., Zhang, L., Yang, M.H.: Real-time compressive tracking. In: Computer Vision–ECCV 2012. Springer (2012) 864–877
23. Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011) 1619–1632
24. Sevilla-Lara, L., Learned-Miller, E.: Distribution fields for tracking. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2012) 1910–1917
25. Kalal, Z., Matas, J., Mikolajczyk, K.: Pn learning: Bootstrapping binary classifiers by structural constraints. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2010) 49–56
26. Jia, X., Lu, H., Yang, M.H.: Visual tracking via adaptive structural local sparse appearance model. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2012) 1822–1829
27. Zhong, W., Lu, H., Yang, M.H.: Robust object tracking via sparsity-based collaborative model. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2012) 1838–1845
28. Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2012) 1830–1837