The 5th Annual IEEE International Conference on
Cyber Technology in Automation, Control and Intelligent Systems
June 8-12, 2015, Shenyang, China

Locality Constrained Low-Rank Sparse Learning for Object Tracking

Baojie fan[1,2]*,Yandong Tang*[2]*,*

1. College of Automation, Nanjing University of Posts and Telecommunications,
Nanjing, 210046, China

2. State Key Laboratory of Robotics, Shenyang Institute Automation, Chinese Academy of Sciences,
Shenyang, 110016*, China*

***Abstract*—In this paper, we present a locality constrained low rank sparse learning algorithm for object tracking under the particle filter framework. Locality should be as important as the sparsity. It can further exploit spatial relationship among particles and increase the consistency of low rank coding. Locality information among the training data and dictionary is mined. This can be achieved by using the local constraints as the regularization term. Combined the low rank and sparse criteria, the total objective function is constructed for locality constrained low rank sparse learning. It can be solved by a sequence of closed form update operations. The best target candidate is chosen by jointly evaluating the reconstructive error and classification error. Extensive experimental results on challenging video sequences demonstrate that the proposed tracking method achieves state-of-the-art performance in term of accuracy and robustness.**

Keywords—object tracking; low rank sparse learning; locality information; collaboration metric

I. INTRODUCTION

The purpose of object tracking is, the goal of visual tracking is as follows: Given the initialized state of a target in the first frame, estimate the states of the tracked target in the sub-sequent frames. Object tracking has a wide range of real-world applications such as intelligent surveillance, robot navigation, autonomous vehicles, traffic monitoring, and so on. Despite that much progress has been made in recent years [6-20], developing a robust tracking algorithm is still a challenging problem due to numerous factors**:** background clutters, large and dynamic appearance changes caused by illumination, rotation, and scaling, abrupt motion, partial or full occlusions and shape deformation.

Sparse representation has been successfully applied to visual tracking under the particle filter framework [9], in which each particle as target candidate is as a sparse linear combination of dictionary atoms that can be dynamically updated to maintain an up-to-date target appearance model. Recently, based on the milestone work, there are some methods that have been developed to improve the L1 tracker in terms of both speed and accuracy [9-17]. Among them, a sparse and low rank tracker is developed in [29-32]. By exploiting the relationship among particles and the temporal consistency of sparse representation, the visual tracking is casted as a low-rank sparse learning problem. But less work focus on the quality of the dictionary. Some tracking algorithms construct the dictionaries naïve. The sampled object or background samples are directly as dictionary items, not selective. The operation makes the dictionary redundant, and ignores the discriminative and structured information from the training data. This maybe result in a decrease of the representation power of the learned dictionary.

Inspired by the above work, in this paper, we present a locality constrained low-rank sparse learning algorithm for object tracking under the particle filter framework. The priori information from training data is exploited effectively to online learn the high quality dictionary and optimal classifier jointly. Specifically, label information from training data is incorporated into the dictionary learning process by adding an ideal-code regularization term. This constraint encourages training data from the same class to have similar re-presentations. By using the local constraints as a regularization term, we consider the local manifold structure between data space and dictionary space. Combined the low rank and sparse criteria, a total objective function is constructed for low rank sparse learning. It can be solved by a sequence of closed form update operations. The best candidate is selected by jointly evaluating the reconstructive error and classification error.

II. RELATED WORK

There is extensive literature on visual object tracking. In this section, we briefly review the most relevant literature to this paper that uses particle filters, low rank sparse learning and dictionary learning. For a more thorough survey of tracking methods, we refer the readers to [1-5]. In general, existing tracking algorithms can be categorized as generative or discriminative. Generative tracking methods learn an appearance model to describe the target object, and then use it to search for the target location that has the most similar appearance to the model. Examples of the generative methods are either templates [7-12] or subspace models [6, 12, 20, 22], fragment-based tracker [6], IVT [7], VTD [8], L1 [9] and its extensions [10-17]. Discriminative methods cast the object tracking task as a binary classification problem that distinguishes the tracked targets from their surrounding backgrounds. Examples of discriminative methods are OAB [18], MIL [19], TLD [20], Struck [21], and FCT [22].

978-1-4799-8730-6/15/$31.00 © 2015 IEEE

Sparse representation has been successfully applied to visual tracking [9]. Its metric is according to finding the best candidate with minimal reconstruction error using target templates and trivial ones. A minimal error bounding strategy is introduced [10] to reduce the number of particles, equal to the number of the L1 norm minimizations for solving. In [11], accelerated proximal gradient (APG) based solution is used to improve the L1 tracker. Liu et al.[12] integrate the dynamic group sparsity into the tracking problem and high dimensional image features are used to improve tracking robustness. Liu et al. [13] also develop a tracking algorithm based on local sparse model which employs histograms of sparse coefficients and the mean shift algorithm for object tracking. However, this method is based on a static local sparse dictionary and may fail when there is similar object in the scenes. Wei et al.[14] propose a robust object tracking algorithm using a collaborative model that combines a sparsity based discriminative classifier and a sparsity-based generative model. Xu et al.[15] develop a simple yet robust tracking method based on the structural local sparse appearance model. Its representation exploits both partial information and spatial information of the tracked target based on a novel alignment pooling method. Inspired by these works, Multi-Task Tracking (MTT) algorithm is develops in [16]. However, the dictionary still include the trivial templates, they will degrade the efficiency and effectiveness of the tracker. Multi-Task Multi-View Tracking (MTMVT) method [17] based on joint sparse representation are developed to exploit the related information shared between particles and views in order to obtain improved performance.

III. TOTAL OBJECTIVE FUNCTION

In our particle filter based tracking method, for t th frame, N particles are randomly sampled around the previous frame state of the tracked object according to a zero-mean Gaussian distribution, and their features (pixel values) are denoted in matrix form as $Y=[y_1,y_2,...,y_n]$, where each column stands for a particle. Each particle can be represented as the linear combination of atoms from the dictionary as $Y=DX+E$, where X is denoted as the representation of the particles with respect to D, and E is the residual term. The dictionary is constructed from the overcomplete sampling sets of the target object and its surrounding background.

$$<D,A,X,E>=\underset{D,A,X,E}{\arg\min}\|X\|_*+\frac{\lambda_1}{2}\sum_{i=1}^{N}\left\|w_i\odot x_{1,i}\right\|_2^2+\lambda_2\|X\|_{1,1}+\lambda_3\|E\|_{1,1} \quad (1)$$

Such that $Y_{new}=D_{new}X+E$, here $Y_{new}=[Y;\lambda_{11}H]$, $D_{new}=[D;\lambda_{11}A]$. For object tracking task, we define two classes: tracked object and background. A linear classifier $f(A;X)=AX$ is learned and updated online, where A are the classifier parameters. $H=[h_1,h_2,...h_N]\in R^{2\times N}$ are the class labels of training data Y. $h_i=[1,0]^t$ is the corresponding label vector of y_i, and the non-zero position indicates the class label of y_i.

To solve Eqn. 1, we rewrite it to the following equivalent format, three equality constraints and auxiliary variables are introduced.

$$\min_{D,A,X_{1-4},E}\|X_1\|_*+\frac{\lambda_1}{2}\sum_{i=1}^{N}\left\|w_i\odot x_{3,i}\right\|_2^2+\lambda_2\|X_2\|_{1,1}+\lambda_3\|E\|_{1,1}$$
$$\text{such that}\begin{cases}Y_{new}=D_{new}X_4+E\\ X_4=X_1\\ X_4=X_2\\ X_4=X_3\end{cases} \quad (2)$$

Using the Inexact Augmented Lagrange Multiplier (IALM) method, we can derive the augmented Lagrangian function of (2) is

$$\begin{aligned}L(D,A,X_{1-4},Y_{1-4},E,\mu_{1-4})=&\|X_1\|_*+\frac{\lambda_1}{2}\sum_{i=1}^{N}\left\|w_i\odot x_{3,i}\right\|_2^2+\lambda_2\|X_2\|_{1,1}+\\&+\lambda_3\|E\|_{1,1}+<Y_1,Y_{new}-D_{new}X_4-E>+\frac{\mu_1}{2}\|Y_{new}-D_{new}X_4-E\|_F^2\\&+<Y_2,X_4-X_1>+\frac{\mu_2}{2}\|X_4-X_1\|_F^2+<Y_3,X_4-X_2>+\frac{\mu_3}{2}\|X_4-X_2\|_F^2\\&+<Y_4,X_4-X_3>+\frac{\mu_4}{2}\|X_4-X_3\|_F^2\end{aligned} \quad (3)$$

where $<Y_2,X_4-X_1>=trace(Y_2^t(X_4-X_1))$, Y_{1-4} are Lagrange multipliers, μ_{1-4} are penalty parameters. The optimization for problem (3) can be divided into two subproblems. The first subproblem is to compute the optimal X_{1-4},E for a given dictionary D_{new}. The second subproblem is to solve dictionary D_{new} for the given X_{1-4},E calculated from the first subproblem.

A. Updating X_{1-4},E Given D_{new}

IALM algorithm is an iterative method in a coordinate descent manner. When each variable updates, other variables are fixed. The updates are in closed form from. So, we can obtain the following update steps corresponding to variables X_{1-4},E

Updating X_1: X_1 can be solved in closed form

$$X_1=\underset{X_1}{\arg\min}\frac{\lambda_1}{\mu_2}\|X_1\|_*+\frac{1}{2}\left\|X_1-X_4-\frac{Y_2}{\mu_2}\right\|_F^2\Rightarrow X_1=J_{\lambda_1/\mu_2}(X_4+\frac{Y_2}{\mu_2}) \quad (4)$$

where $J_\lambda(P)=US_\lambda(\Sigma)V^T$, $P=U\Sigma V^T$, and $S_\lambda(P_{i,j})$ is the soft-thresholding operator, the traditional solver is as $S_\lambda(P_{i,j})=sign(P_{i,j})\max(0,|P_{i,j}|-\lambda)$. In this paper, a generalized iterated shrinkage algorithm in [38]is used to obtain the stable and efficient.

Updating X_2: X_2 is computed with the closed form solution from the equation (5).

$$X_2=\underset{X_2}{\arg\min}\frac{\lambda_2}{\mu_3}\|X_1\|_{1,1}+\frac{1}{2}\left\|X_2-X_4-\frac{Y_3}{\mu_3}\right\|_F^2\Rightarrow X_2=S_{\lambda_2/\mu_3}(X_4+\frac{Y_3}{\mu_3}) \quad (5)$$

Updating X_3: X_3 is updated by solving the optimization problem (6) with the closed form solution.

$$X_3 = \arg\min_{X_3} \|X_3\|_F^2 + \frac{\lambda_1}{2}\sum_{i=1}^{N}\|w_i \odot x_{3,i}\|_2^2 + \langle Y_4, X_4 - X_3\rangle + \frac{\mu_4}{2}\|X_4 - X_3\|_F^2 \quad (6)$$

$$X_{3,i} = \arg\min_{X_{3,i}} \frac{1}{2}\sum_{i=1}^{N}\lambda_3\|x_{3,i}\|_2^2 + \lambda_4\|w_i^T x_{3,i}\|_2^2 + \mu_4\|x_{4,i} - x_{3,i}\|_2^2 + 2Y_{4,i}^T(x_{4,i} - x_{3,i}) = (\lambda_4 w_i^T w_i + \lambda_3 I + \mu_4 I)^{-1}[-\mu_4 x_{4,i} - Y_{4,i}] \quad (7)$$

where $x_{3,i}$ is the i-th item of X_3 as $X_3 = [x_{3,1}, x_{3,2}, ...x_{3,N}]$

Updating E : E can be computed with the closed form solution from the equation (8).

$$E = \arg\min_E \frac{\lambda_5}{\mu_1}\|E\|_{1,1} + \frac{1}{2}\left\|E - D_{new}X_4 + Y_{new} - \frac{Y_1}{\mu_1}\right\|_F^2 \Rightarrow E = S_{\lambda_5/\mu_1}(D_{new}X_4 - Y_{new} + \frac{Y_1}{\mu_1}) \quad (8)$$

Updating X_4 : X_4 can be computed with the closed form solution from the equation (9)

$$X_4 = \arg\min_{X_4} \langle Y_1, Y_{new} - D_{new}X_4 - E\rangle + \frac{\mu_1}{2}\|Y_{new} - D_{new}X_4 - E\|_F^2 + \langle Y_2, X_4 - X_1\rangle + \frac{\mu_2}{2}\|X_4 - X_1\|_F^2 + \langle Y_3, X_4 - X_2\rangle + \frac{\mu_3}{2}\|X_4 - X_2\|_F^2 + \langle Y_4, X_4 - X_3\rangle + \frac{\mu_4}{2}\|X_4 - X_3\|_F^2 \quad (9)$$

$$\Rightarrow X_4 = G_1[D_{new}^T(Y_{new} - E) + G2 + G3]$$

where $G1 = (D_{new}^T D_{new} + \frac{\mu_2 + \mu_3 + \mu_4}{\mu_1} I)^{-1}$, $G2 = \frac{\mu_2 X_1 + \mu_3 X_2 - \mu_4 X_3}{\mu_1}$,

$$G3 = \frac{1}{\mu_1}(D_{new}^T Y_1 - Y_2 - Y_3 - Y_4) .$$

Updating Y_{1-4} , μ_{1-4} :

$$\begin{cases} Y_1 = Y_1 + \mu_1(Y_{new} - D_{new}X_4 - E) \\ Y_2 = Y_2 + \mu_2(X_4 - X_1) \\ Y_3 = Y_3 + \mu_3(X_4 - X_2) \\ Y_4 = Y_4 + \mu_4(X_4 - X_3) \\ \mu_i = \max(1.1 \times \mu_i, 10^6) \end{cases} \quad (10)$$

B. Updating D_{new} Given X_{1-4}, E

With the given X_{1-4}, E , D_{new} is the only variable in second subproblem. So equation (4) can be rewrited as the following formula by retaining all the variables associated with D_{new} .

$$D_{new} = \arg\min_{D_{new}} \langle Y_1, Y_{new} - D_{new}X_4 - E\rangle + \frac{\mu_1}{2}\|Y_{new} - D_{new}X_4 - E\|_F^2 + \frac{\lambda_4}{2}\sum_{i=1}^{N}\|w_i \odot x_{3,i}\|_2^2 \quad (11)$$

In [33], when optimizing the above equation, the last term is neglected to obtain

$$D_{new} = \arg\min_{D_{new}} \langle Y_1, Y_{new} - D_{new}X_4 - E\rangle + \frac{\mu_1}{2}\|Y_{new} - D_{new}X_4 - E\|_F^2 \quad (12)$$

By using gradient descent, we can solve the above problem

$$D_{new}^{i+1} = D_{new}^i + \alpha\nabla D_{new}^i \quad (13)$$

where $D_{new} = \frac{1}{\mu_1}(Y_1 + \mu_1(Y_{new} - E))X_4^T(X_4X_4^T)^{-1}$, α is the step length that controls the learning rate.

C. Collaboration Metric Scheme

Then, combined the online learned dictionary and robust sparse representation, the joint decision measure is developed to evaluate the reliability of candidates. The trace is completed by the following equation:

$$\bar{X}_i = \arg\min_{X_i}(\|Y - D^*X\|^2 + \delta\|H - A^*X\|^2) \quad (14)$$

where $\delta = 0.5$ is the weight of the classifier. The best candidate particle is selected by jointly evaluating the reconstruction error and classification error.

IV. EXPERIMENTS

In this section, we demonstrate the merits of the proposed algorithm with extensive experimental results. The thorough comparisons are done on challenging image sequences between our proposed trackers and state of the art tracking methods.

A. Datasets and Baseline

The proposed tracker is evaluated on 8 challenging tracking sequences that are publicly available online. These videos contain various challenges, such as motion blur, the changes of pose, illumination, rotation and scale, partial and heavy occlusion, background clutter.

We evaluate the proposed tracker against fifteen state-of-the-art visual tracking algorithms including: ONND [35], CN [36], LRST[30], SCM [14], Struck [21], CT [22], ASLA [15], MTT [16], VTD [8], MIL [19], PN [20], IVT [7], Frag [6] and L1 [9], et al. These trackers are implemented using publicly available source codes or binaries provided by the authors. They are initialized using their default parameters.

B. Implementation Details

The proposed algorithm is implemented in MATLAB R2011b on a Pentium 2.3 GHz Dual Core laptop with 2GB RAM. Each image sample from the target and background is normalized to a 32 × 32 patch and represented by 1024-dimensional vector of intensity values. The numbers of object samples and background samples are 40 and 100 respectively. The parameters $\lambda_1, \lambda_2, \lambda_3$ of the proposed algorithm are set as 2.0, 2.0, 0.1 and 1.0 respectively. In addition, we set $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

C. Quantitative Comparison

Two popular evaluation criteria were used to quantitatively evaluate the 16 trackers, namely, center location error (CLE) and tracking success rate (TSR). The CLE was computed as the distance between the tracked object center position and the ground truth center position. Table 1 summarizes the average center location errors in pixels. The TSR was computed as the ratio of the number of frames the target was successfully tracked to the number of frames in the sequence. To define whether the target is successfully tracked at a frame, we use the score in the PASCAL VOC challenge [37], which can be computed as

$$score = {area(R_T \cap R_G)} \Big/ {area(R_T \cup R_G)} \qquad (15)$$

where R_T is the current the tracking result and R_G is the ground truth. if $score > 0.5$, the result of one frame is considered as a success. Table 2 gives the average success rates.

From the quantitative results in Table 1 and Table 2, we can see that proposed tracker achieves the best or second best performance in many tested sequences both in terms of CLE and TSR. Overall, the proposed tracker performs well against the other state-of-the-art algorithms.

TABLE I. The average center location errors of 16 trackers on 8 sequences. The best three results are shown in red, blue, and green fonts

	animal	jump	david11	Car11	trellis	cliff	football	girl
IVT	127.5	36.80	3.589	2.106	107.9	24.82	36.95	47.90
frag	92.09	58.45	76.69	63.92	56.32	48.67	17.21	18.06
MIL	66.46	9.894	16.15	43.47	68.78	13.35	13.66	32.29
L1	171.5	92.39	7.630	33.25	34.32	49.60	18.17	60.98
PN	25.65	3.589	9.671	25.11	40.30	11.25	13.543	23.23
VTD	11.92	62.99	13.55	27.05	35.38	34.56	13.61	21.48
MTT	15.86	34.47	143.29	2.802	48.65	46.17	9.842	23.88
ONN	8.443	36.61	23.48	1.782	8.594	29.61	20.37	27.75
SCM	10.02	4.097	5.344	1.686	12.83	7.706	3.900	9.934
ASLA	7.174	4.280	3.658	1.582	49.51	5.606	14.94	16.08
LSST	10.01	4.772	3.937	1.601	85.70	23.31	7.574	72..34
Struck	13.55	7.292	17.62	4.513	23.34	7.998	13.93	128.7
CT	19.85	43.00	16.47	8.352	41.20	23.42	8.138	32.81
CN	9.066	38.48	15.52	1.617	21.75	3.309	16.82	31.91
LRST	43.71	22.90	16.27	4.148	11.75	20.25	4.396	18.75
Our	8.212	4.036	3.628	1.464	7.695	3.7785	3.831	10.27

TABLE II. The average tracking success rate of of 16 trackers on 8 sequences. The best three results are shown in red, blue, and green fonts

	animal	jump	david1	car11	trellis	cliffbar	football	girl
IVT	0.2166	0.2826	0.7116	0.8077	0.0759	0.5648	0.1269	0.4305
frag	0.0764	0.1383	0.1946	0.0857	0.3176	0.1337	0.5210	0.6888
MIL	0.2129	0.5267	0.4479	0.1745	0.2666	0.4622	0.5760	0.5189
L1	0.0386	0.0927	0.6250	0.4353	0.3901	0.1993	0.5732	0.3296
PN	0.4118	0.6904	0.6015	0.3761	0.1473	0.3798	0.5049	0.5759
VTD	0.5771	0.0797	0.5252	0.4320	0.3986	0.3292	0.5573	0.5110
MTT	0.5185	0.2318	0.2888	0.7537	0.4168	0.3073	0.6643	0.6338
ONND	0.6257	0.1357	0.4131	0.7930	0.6929	0.3483	0.4076	0.4218
SCM	0.6081	0.7174	0.6712	0.7981	0.6516	0.6533	0.8308	0.6795
ASLA	0.6198	0.7121	0.6430	0.8425	0.1462	0.6197	0.6362	0.6505
LSST	0.5750	0.6540	0.7671	0.8376	0.2953	0.5648	0.6886	0.1211
Struck	0.6062	0.5819	0.4442	0.6589	0.4503	0.6580	0.5257	0.0103
CT	0.5250	0.1531	0.4342	0.5306	0.3413	0.3852	0.6994	0.5124
CN	0.6039	0.1242	0.4522	0.8245	0.4585	0.6103	0.6076	0.5547
LRST	0.3185	0.2970	0.4902	0.6976	0.6127	0.4480	0.7978	0.6451
Our	0.6300	0.7049	0.7731	0.8380	0.7215	0.6361	0.8016	0.8085

D. Quantitative Comparison

The qualitative comparison is shown in figure 1. Due to the limited space available, we try our best to illustrate the tracking results of different trackers in some key frames from each tested sequence.

In animal sequence, fast motion of the target object leads to many blurred images. From 23 to 50, the head of the fawn in image sequence becomes blurred. The trackers such as MIL, PN, CT, Frag, LRST, MTT fail to follow the target. The proposed algorithms can overcome this difficulty, and successfully locate the target object throughout the sequence. The accuracy and overlap rate are better than LSRT, VTD, CN, SCM, LSST, less than ASLA, ONND.

As the man jumps up and down, it is rather challenging to account for drastic appearance change caused by motion blur. Figure 1(b) shows that when the man skips, he move so fast, some tracking algorithms fail to follow the target right at the beginning section of this sequence (e.g. #20, 40). Our tracker has good performance in this sequence.

In david1 sequence, the appearance of the person changes gradually due to illumination and pose variation when he walks from a dark room into areas with spot light. The IVT, SCM, ASLA, LSST and our trackers track the moving face accurately, as shown in figure 1. (c). our tracker has excellent performance with the highest locating accuracy. Some other trackers cannot complete the trace, especially when the person is with the out-of-plane rotation around the frame 160.

The tracking results of the car11 sequence are illustrated in Figure 1(d). CN, LSRT, IVT, SCM, ASLA, LSST, MTT and ONND algorithms perform well as our trackers in the whole sequence. However, the accuracy and robustness of these methods are less than LSST and ASLA algorithms. But other methods drift away when drastic illumination variation occurs from the 240 frame to 290 frame or when similar objects appear in the scene (e.g., #0305); especially the car makes a turn at about frame 262.

In trellis sequence, a person walks underneath a trellis covered in vines. The cast shadow changes the appearance of the target face significantly. In figure 3. (e). SCM, ONNDL and our trackers have good performance and can locate the target accurately and robustly. Our-L2 has the highest accuracy. In contrast, the other tracking methods perform poor, especially when the person is with the combined effects of pose and lighting variations.

In the cliffbar video, the target undergoes background clutter, scale variance, in-plane rotation, and abrupt motion as shown in figure 1(f). The Frag, L1, IVT, CT, MIL, LSST, ONND, SCM tracking methods drift to the similar background, among them, some trackers do not exploit the background information and lack the discriminative ability to keep track of the objects correctly. Our proposed trackers have the excellent performance on this sequence with the smallest tracking errors.

There are many football players with the similar helmets in appearance to the tracked object in this scene. When the two football players collide at frame 290, most tracking methods cannot locate the target correctly. Only our tracker, CT, VTD, and ONND overcomes this problem and successfully locate the correct object in the whole sequence. The accuracy of our method is the highest.

In the girl sequence, the tracking object undergoes occlusion, large pose change, and scale variation with in-plane and out-of-plane rotations. As shown in Fig. 1(h). The experimental results demonstrate that our proposed method achieves the best performance in this sequence, owing to the label information and combined metric.

The tracking results on the sylv sequence are shown in Fig. 1(i). In this sequence, a stuffed animal is being moved, thus, leading to challenging pose, lighting, and scale changes, rotations in plane or out of plane. The MIL, TLD, Struck, CN, SCM, ONNDL and our proposed trackers perform well on this sequence with lower tracking errors than other methods. L1, IVT, ASLA, LSST methods do not perform well on this sequence as these methods use holistic features which are less effective for large scale pose variations.

(a) animal

(b) jumping

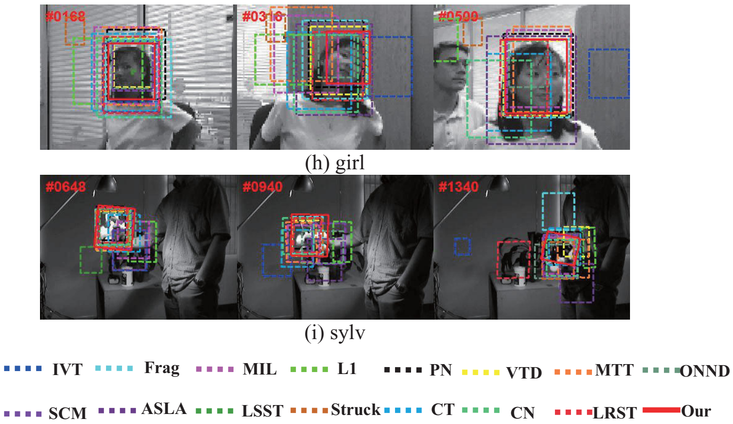
(c) david1

(d) car11

(e)trellis

(f) cliffbar

(g) football



(h) girl

(i) sylv

Figure 1: Comparisons of tracking results of 16 trackers (denoted in different colors) on 8 video sequences.

V. CONCLUSIONS

In this paper, a locality constrained low-rank sparse learning algorithm is proposed for robust object tracking under the framework of particle filter. A total objective function is constructed for low rank sparse learning. Which have the locality constrained property and low-rank sparse property,. Compared with traditional low rank and sparsity of the representation, we should focus on the locality information of the training data and dictionary templates. Meanwhile, label information from training data is incorporated into the dictionary learning process to online learn the high quality dictionary and optimal classifier jointly. Both quantitative and qualitative comparisons demonstrate that the proposed algorithms perform well in terms of accuracy and robustness. Future work contains incorporating the low rank feature selection scheme and the efficient optimizing method.

Acknowledgment

This work has been funded by the Natural Science Foundation of China (grant no. 61203270), State Key Laboratory of Robotics Open Project, the Introduction Foundation for the Talent of Nanjing University of Tele. and Com(no. NY212028) and the Jiangsu Key Laboratory of Image and Video Understanding for Social Safety (Nanjing University of Science and Technology), Grant No. 30920140122007.

References

[1] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, *ACM Computing Surveys*, 38(4), 2006.

[2] X. Li, W. Hu, C.Shen, Z.Zhang, A.Dick, A.van den Hengel, A survey of appearance model in visual object tracking, *ACM Transactions on Intelligent Systems and Technology*, 2013.

[3] Y. Wu, Jongwoo Lim, and Ming-Hsuan yang, Online object tracking: a Benchmark, *In CVPR*, 2013.

[4] S., Zhang, H., Yao, Xin Sun, X., Lu, Sparse coding based visual tracking: Review and experimental comparison, *Pattern Recognition*, 46(7): 1772--1788, 2012.

[5] A. Smeulder, D. Chu, R. Cucchiara, S. Calderara, A. Deghan, M. Shah, Visual Tracking: an Experimental Survey, *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

[6] A. Adam, E. Rivlin, I. Shimshoni, Robust fragments-based tracking using the integral histogra. *In CVPR*, pp: 798--805, 2006.

[7] D. Ross, J. Lim, R. Lin, M. Yang, Incremental learning for robust visual tracking, *International Journal of Computer Vision*, 77(1): 125–141, 2008.

[8] J., Kwon, K. M. Lee, Visual tracking decomposition. *In CVPR*, pp:1269–1276, 2010.

[9] X. Mei and H. Ling. Robust visual tracking and vehicle classification via sparse representation. *TPAMI*, 33(11): 2259–2272, 2011.

[10] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai. Minimum error bounded efficient l1 tracker with occlusion detection. *In CVPR*, pp: 1257--1264, 2011.

[11] C. Bao, Y.Wu, H. Ling, and H. Ji. Real time robust ℓ1 tracker using accelerated proximal gradient approach. *In CVPR*, pp:1830--1837, 2012.

[12] B. Liu, J. Huang, L. Yang, & C. Kulikowski. Robust visual trackingwith local sparse appearance model and k-selection. *In CVPR*, pp:1--8, 2011.

[13] B., Liu, L., Yang, J., Huang, P., Meer, L., Gong, & C. Kulikowski. Robust and fast collaborative tracking with two stage sparse optimization. *In ECCV*, pp:1--14, 2011.

[14] W. Zhong, H. Lu, and M.-H. Yang. Robust object tracking via sparsity-based collaborative model. *In TIP*, 23(5): 2356--2368, 2014.

[15] X. Jia, H. Lu, and M.-H. Yang. Visual tracking via adaptive structural local sparse appearance model. *In CVPR*, pp: 1822--1829, 2012.

[16] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Robust visual tracking via structured multi-task sparse learning. *International Journal of Computer Vision*, 101(2): 367--383, 2013.

[17] Z. B. Hong, X. Mei, D. Prokhorov, D. C. Tao, Tracking via Robust Multi-Task Multi-View Joint Sparse Representation, *In ICCV2013*, pp: 649--656, 2013.

[18] H. Grabner, M. Grabner, H. Bischof, Real-time tracking via online boosting. *In BMVC*, pp: 47--56, 2006.

[19] B. Babenko, M.-H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1619--1632, 2011.

[20] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking learning detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1409--1422, 2012.

[21] S. Hare, A. Saffari, and P. H. Torr. Struck: Structured output tracking with kernels. In *ICCV*, pp: 263--270, 2011.

[22] K. Zhang, L. Zhang, and M-H. Yang., Fast compressive tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10): 2002--2015, 2014.

[23] G. Liu and S. Yan. Latent low-rank representation for subspace segmentation and feature extraction. *In ICCV*, 2011.

[24] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. on PAMI*, 35(1):171--184, 2013..

[25] J. Lee, B. Shi, Y.Matsushita, I. Kweon, and K. Ikeuchi. Radiometric calibration by transform invariant low-rank structure. *In CVPR,* 2011.

[26] B. Cheng, G. Liu, J.Wang, Z. Huang, and S. Yan. Multi-task low-rank affinities pursuit for image segmentation. *In ICCV*, 2011.

[27] X. Shen and Y.Wu. A unified approach to salient object detection via low rank matrix recovery. *In CVPR*, 2012.

[28] Y., Peng, A., Ganesh, J.,Wright, W., Xu, & Y. Ma, RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(11), 2233–2246, 2011.

[29] X. Cui, J. Huang, S. Zhang, and D. Metaxas. Background subtraction using group sparsity and low rank constraint. *In ECCV*, 2012.

[30] T., Zhang, B., Ghanem, S., Liu, & N. Ahuja, Low-rank sparse learning for robust visual tracking. *In ECCV, 2012.*

[31] T., Zhang, C., Jia, C., Xu, Y., Ma, & N. Ahuja, Partial occlusion handling for visual tracking via robust part matching. In *CVPR, 2014*.

[32] T., Zhang, S., Liu, N. Ahuja, M-H. Yang, B. Ghanem, Robust Visual Tracking via Consistent Low-Rank Sparse Learning, *International Journal of Computer Vision*, 2014.

[33] J., Wang, J., Yang, K., Yu, F., Lv, T., Huang, Y. Gong, Locality constrained linear coding for image classification. *In CVPR*, pp: 3360--3367. 2010.

[34] E. J., Candes, Y. Plan, Matrix completion with noise.Proceedings of the IEEE 98(6): 925–936, 2010.

[35] N. Wang, J. Wang, D. Yeung. Online robust non-negative dictionary learning for visual tracking. *In ICCV*, 2013.

[36] M. Danelljan, F. S. Khan, M. Felsberg and J. Weijer. Adaptive color atributes for real-time visual tracking. *In CVPR*, 2014.

[37] M. Everingham, L. Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object class (voc) challenge, *Int. J. Comput. Vision*, 88: 303--338, 2010.

[38] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, A generalized iterated shrinkage algorithm for non-convex sparse coding, In ICCV 2013.