ONLINE DISCRIMINATIVE DICTIONARY LEARNING VIA LABEL INFORMATION FOR MULTI TASK OBJECT TRACKING

Baojie Fan[1]*, *Yingkui Du*[2], *Hao Gao*[1], *Baoyun Wang*[1]
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
E-mail: jobfbj@gmail.com

ABSTRACT

In this paper, a supervised approach to online learn a structured sparse and discriminative representation for object tracking is presented. Label information from training data is incorporated into the dictionary learning process to construct a compact and discriminative dictionary. This is accomplished by adding an ideal-code regularization term and classification error term to the total objective function. By minimizing the total objective function, we learn the high quality dictionary and optimal linear multi-classifier simultaneously. Combined with multi task sparse learning, the learned classifier is employed directly to separate the object from background. As the tracking continues, the proposed algorithm alternates between multi task sparse coding and dictionary updating. Experimental evaluations on the challenging sequences show that the proposed algorithm performs favorably against state-of-the-art methods in terms of effectiveness, accuracy and robustness.

***Index Terms*—**label information, discriminative dictionary learning, multi task learning, object tracking

1. INTRODUCTION

Visual tracking is a well-known topic in computer vision with many applications such as automated surveillance, robot navigation, medical imaging, video indexing, traffic monitoring, human-computer interaction, etc. It is a challenging task to develop a robust tracking algorithm due to numerous factors: illumination, partial or full occlusions, dynamic appearance changes, scaling, abrupt motion, background clutters, pose variation and shape deformation. To overcome these difficulties, much progress has been made in recent years [1-20]. To survey many of these algorithms, we refer the reader to [1-4].

Mei and Ling [7] proposed the L1 tracker (L1T) for robust visual tracking under the particle filter framework based on the sparse coding technique. In detail, the target templates are used to describe the tracked object and trivial templates are used to deal with outliers. This representative scheme and its extensions are robust to a wide range of image corruptions, especially moderate occlusions. But, less work focus on the quality of the dictionary during the tracking process.

In this paper, we formulate object tracking in a particle filter framework as a binary classification problem. The priori information from training data is exploited effectively to online learn a discriminative and reconstructive dictionary. Specifically, the class label information is incorporated into the dictionary learning process as the classification error term and idea coding regularization term respectively. Combined with the traditional reconstruction error, a total objective function for dictionary learning is constructed. By minimizing the total object function, we can obtain a high quality dictionary and optimal linear classifier jointly. Combined with the multi task sparse coding, the optimal classifier can separate the tracker object from background effectively.

The main contributions of this paper are:

(1) The priori information from the training samples is exploited to construct a compact and discriminative dictionary. It is a critical factor for the object tracker based sparse representation. The learned dictionary encourages samples from the same class to have similar representations.

(2) Learning a high quality dictionary and optimal linear classifier are accomplished jointly. All the training samples from the object and background are involved the dictionary learning process simultaneously.

(3) Experiments show the proposed tracker outperforms some state-of-the-art methods on challenging sequences with heavy occlusion, drastic illumination changes, and large pose variations.

2. RELATED WORK

In this section, we briefly review nominal tracking methods and those that are the most related to our tracker. We focus specifically on the representative tracking methods that use particle filters, sparse representation and dictionary learning. For a more thorough survey of tracking methods, we refer the readers to [1-4]. Existing tracking algorithms can be roughly categorized as either generative or discriminative.

Generative tracking methods learn an appearance model to describe the target observations, and the aim is to search for the target location that has the most similar appearance to the model. Examples of generative methods are IVT [5], VTD [6], L1 [6-9], SPT [10], SCM [11], and MTT [15].

Ross et al. [5] learn an adaptive linear subspace online for modeling target appearance and implement tracking with a particle filter. Kwon et al. [6] extend the classic particle filter framework with multiple dynamic observation models to account for appearance and motion variation. L1 [7] is the most representative work, and some extensions [8-13] are developed to improve the L1 tracker in terms of both speed and accuracy. In [9], APG based solution is used to improve the L1 tracker. Liu et al. [10] also develop a tracking algorithm based on local sparse model which employs histograms of sparse coefficients and the mean shift algorithm for object tracking. In Zhang et al. [12], low-rank sparse learning is adopted to consider the correlations among particles for robust tracking. Inspired by these works, he develops the Multi-Task Tracking (MTT) algorithm [13]. However, the dictionary still include the trivial templates, they will degrade the efficiency and effectiveness of the tracker. In Wang et al [14], online robust non-negative dictionary learning method is developed for visual tracking, a new particle representation formulation using the Huber loss function is proposed to estimate the robust object templates.

Discriminative methods cast the tracking as a binary classification problem that distinguishes the tracked targets from their surrounding backgrounds. The trained classifier is online updated during the tracking procedure. Examples of discriminative methods are MIL [15], PN [16], CT [17] and Struck [18].

Babenko et al. [15] introduce multiple instance learning into online tracking where samples are considered within positive and negative bags or sets. Kalal et al. [16] propose the PN learning algorithm to exploit the underlying structure of positive and negative samples to learn effective classifiers for object tracking. The Struck [17] ranks top in the recent benchmark [3], and it learns a kernelized structured output support vector machine online. An efficient tracking algorithm [18] based on compressive sensing theory [19] is proposed, which demonstrates that the low dimensional features randomly extracted from the high dimensional multi-scale image feature space can preserve the discriminative capability, thereby facilitating object tracking.

3. BACKGROUND

In this section, we briefly introduce the particle filter and dictionary learning to facilitate the presentation of our model in the next section.

3.1. Particle filter

Particle filters is a popular tracking framework due to its excellent performance in the presence of nonlinear target motion and to flexibility to different object representations. It can be considered as a Bayesian inference task in a Markov model with hidden state variables, which recursively approximates the posterior distribution using a finite set of weighted samples. It consists of two steps: prediction and update.

Specially, at the frame, let affine parameters $X=(x,y,s,r,\theta,\lambda)$ represent the target state, where x and y are the image coordinates, s and r are the scale and the aspect, θ is the rotation angle, λ is the skew. $Y_{1:t-1}=\{Y_1,Y_2,\ldots,Y_{t-1}\}$ denotes the observation of the target from the first frame to the frame $t-1$. Particle filters tracking estimates and propagates the probability by recursively performing prediction

$$p(X_t \mid Y_{1:t-1}) = \int p(X_t \mid X_{t-1}) p(X_{t-1} \mid Y_{1:t-1}) dX_{t-1} \tag{1}$$

and updating

$$p(X_t \mid Y_{1:t}) = \frac{p(Y_t \mid X_t) p(X_t \mid Y_{1:t-1})}{p(Y_t \mid Y_{1:t-1})} \quad (2).$$

The optimal state for the frame t is obtained according to the maximal approximate posterior probability

$$X_t^* = \arg\max_X p(X \mid Y_{1:t}) \tag{3}$$

This inference is governed by the model $p(X_t \mid X_{t-1})$, which describes the temporal correlation of the tracking results in consecutive frames, and it is modeled to be Gaussian with the dimensions of X_t assumed independent. The observation model $p(Y_t \mid X_t)$ reflects the similarity between a target candidate and dictionary templates. In this paper, $p(Y_t \mid X_t)$ is proportional to the classifier scores.

3.2. Dictionary learning

The goal of dictionary learning is to find the optimized dictionaries that provide a succinct representation for most statistically representative input signals.

Let $Y=[y_1,y_2,\ldots y_N]\in R^{n\times N}$ be a set of N input signals, where y_i denotes the i-th input signal with n dimensional feature description. Learning a reconstructive dictionary with size K for sparse representation can be obtained by solving the following minimization problem

$$< D^*, X^* >= \arg\min_{D,X} \sum_{i=1}^{N} \left(\|y_i - Dx_i\|_2^2 \right) + \lambda \|x_i\|_1) \quad (4).$$

where $D=[d_1,d_2,\ldots d_K]\in R^{n\times K}$ is the learned dictionary, $X=[x_1,x_2,\ldots x_N]\in R^{K\times N}$ is the sparse codes of input signals. In general, the number of training samples is larger than the size of $D(N>>K)$, and x_i only uses a few dictionary atoms for its representation under the sparsity constraint. Usually, the above objective function is iteratively optimized in a two stage manner, by alternatively optimizing with respect to D (bases) and X (coefficients) while holding the other fixed. The objective function in (4) only focuses on minimizing the reconstruction error and does not consider the discriminative power of a dictionary.

4. JOINTLY LEARNING DISCRIMINATIVE DICTIONARY AND OPTIMAL CLASSIFIER

Inspired by the above work, an approach to online learn a structured sparse and discriminative representation for object tracking is presented in this section. We develop a supervised learning method to construct a discriminative and reconstructive dictionary. The class label and structure information among samples are incorporated into the dictionary learning process as the discriminative term and structured regularization term respectively. Combined with the traditional reconstruction error term, a unified objective function for object tracking can be obtained. In this way, the dictionary and the classifier are learned jointly. With the high quality dictionary, structured sparsity based discriminative classifier can be directly used for object tracking.

4.1 The unified object function

To be concrete, the objective function for our object tracking is defined as

$$\begin{aligned} < D^*, A^*, X^* > = \arg\min_{D,A,X} \|Y - DX\|_2^2 \\ + \lambda_1 \|H - AX\|_2^2 + \lambda_2 \|Q - X\|_2^2 + \lambda_3 \|X\|_{2,1} \end{aligned} \quad (5).$$

where parameters $\lambda_1, \lambda_2, \lambda_3$ control the relative weight of three terms: reconstruction error term, classification error, idea coding regularization term, mixed norm regularization term.

Reconstruction Error term $\|Y - DX\|_2^2$: This data fitting term is widely used in sparse representation based tracking [7] and dictionary learning. Its value reflects the presence of occlusion and whether a candidate particle is sampled from the background. We compute the reconstruction errors of all the particles with the learned dictionary items at the same time.

Ideal structured regularization term $\|Q - X\|_2^2$ **:** This term includes the structured and discriminative information from training samples. $Q = [q_i, q_2, \ldots q_M] \in R^{K \times M}$ is the idea representation for Y , M is the number of training samples. We hope that X is very close to Q , and force the samples from the same class to have similar discriminative sparse representation without losing structure information. q_i is the sparse code corresponding to an input signal y_i with the form $q_i = [q_i^1, q_i^2, \ldots q_i^K]' = [0, \ldots 1,1,1,\ldots]' \in R^K$. We cast the object tracking can be viewed as a binary classification problem: object (class T) and background (class B). If the training data is sampled from the tracked object region, the coefficients in q_i for class T are all 1s, while the others are all 0s. For example, the training samples $Y = [y_1, y_2, y_3, y_4]$ include two classes: y_1, y_2 belong to object T and y_3, y_4 are from background B , the ideal representation Q for Y is as follows:

$$Y = DQ = D * \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad (6)$$

Classification error term $\|H - AX\|_2^2$ **:** the term measures the classification error, and it supports learning an optimal classifier. A simple linear classifier $f(A; X) = AX$ is adopted, where $H = [h_1, h_2, \ldots h_N] \in R^{2 \times N}$ are the class labels of training data Y . A is the classifier parameters. $h_i = [1,0]^t$ is the corresponding label vector of y_i , and the non-zero position indicates the class label of y_i .

L1 norm regularization term $\|X\|_{21}$ **:** By adding a sparseness criterion into the objective function (5), we are able to learn a sparse and structural representation with the learned high-quality dictionary D_t . The proposed tracker is under the particle filter framework. The candidate particles are densely sampled around the current tracking target and their representations will be sparse and similar with respect to the given dictionary D_t . In other words, a few items in D_t are required to represent all the particles.

4.2 Optimization procedure

To solve optimization problem in the equation (5), the proposed algorithm alternates between robust sparse coding and dictionary updating. We rewrite the proposed object function as two steps.

Dictionary learning:

$$< D^*, A^* > = \arg\min_{D,A} \|Y_{L1} - D_{L1}X\|_2^2 + \lambda_2 \|Q - X\|_2^2 + \lambda_3 \|X\|_{2,1} \quad (7)$$

where $Y_{L1} = [Y, \sqrt{\lambda_1} H]'$, $D_{L1} = [D, \sqrt{\lambda_1} A]'$.

Sparse coding:

$$X^* = \arg\min_{X} \|Y_{L2} - D_{L2}X\|_2^2 + \lambda_3 \|X\|_{2,1} \quad (8)$$

where $Y_{L2} = [Y, \sqrt{\lambda_1} Q, \sqrt{\lambda_2} H]'$ $D_{L2} = [D, \sqrt{\lambda_1} I, \sqrt{\lambda_2} A]'$.

For equation (7), this is the classical dictionary learning problem that K-SVD [20-21] and online dictionary learning for sparse coding [22] can both obtain the satisfied solution. We combine the methods in [20] and [22] to solve the formula (4) to online learn and update the high quality dictionary. In addition, samples from the same class have low rank sparse codes, which is very important to separate the object from background with the optimal classifier in object tracking process.

Let $E_i = Y_L - \sum_{j \neq i} d_{new}^j x_j^R$, where d_{new}^j is the dictionary item and x_j^R is the *j-th* row in its corresponding coefficients. Discarding the zero atoms in E_i and x_j^R , we are able to obtain $\bar{E}_i$ and $\bar{x}_j^R$ respectively. $d_i, \bar{x}_j^R$ are computed by minimizing the following object function:

$$< d_i, \bar{x}_j^R > = \arg\min_{d_i, \bar{x}_j^R} \left\| \bar{E}_i - d_i \bar{x}_j^R \right\|_F^2 \quad (9).$$

By performing the SVD operation $U\Sigma V^{'} = SVD(\bar{E}_i)$, the solution of equation (5) is as follows:

$$d_i = U(:,1), \ \bar{x}_j^R = \Sigma(1,1)V(:,1) \quad (10).$$

Initialization: This process is completely supervised. The training data set Y is composed of object samples and background samples. Given the initial dictionary D_0, the sparse coding X_0 of the training data Y can be obtained. Then, we employ the multivariate ridge regression model to compute the initial A_0 for training data.

$$A_0 = \arg\min_A \|H - AX_0\|^2 + \gamma\|A\|_2^2 \Rightarrow A_0 = HX_0^{'}(X_0 X_0^{'} + \gamma I)^{-1} \quad (11)$$

Online sparse representation: In frame t, with the tracking result in previous frame, candidate particles Y are densely sampled at a small distance around the target as. With the similar operation, we sample the object samples and background samples as the labeled training data. Then, we construct the corresponding class label vector h_i, and idea representative vector q_i. Given the learned dictionary D_{t-1}, we can compute the corresponding sparse representation for training data by solving the following the optimization problem:

$$X = \arg\min_X \|Y - DX\|_2^2 + \lambda_3 \|X\|_{2,1} \quad (12)$$

We develop the method in [13] and format the object tracking as multi-task sparse learning problem with the high quality dictionary. Learning the representation of each particle is a task, each task is viewed as *jointly sparse*. The joint sparsity is achieved by imposing the $L_{p,q}$ mixed-norm penalty on the reconstruction coefficients. We use the popularly applied Accelerated Proximal Gradient method to efficiently solve the minimization problem in equation (12).

5. EXPERIMENTS

In this section, we demonstrate the merits of the proposed algorithm with extensive experimental results. Our trackers are evaluated on 9 challenging tracking sequences (e.g. car11, cliffbar, singer1 and woman sequences) that are publicly available online. We evaluate the proposed tracker against ten state-of-the-art visual tracking algorithms including: ONND [14] LSST [8], MTT [13], CT [18], VTD [6], MIL[15], PN[16], IVT[5], and L1[7]. These trackers are implemented using publicly available source codes or binaries provided by the authors. They are initialized using their default parameters.

5.1 Parameter setting

The proposed algorithm is implemented in MATLAB R2011b on a Pentium 2.3 GHz Dual Core laptop with 2GB memory. For each sequence, the location of the target object is manually labeled in the first frame. Each image sample from the target and background is normalized to a 32 × 32 or 48 × 16 patch. We set the parameters $\lambda_1, \lambda_2, \lambda_3$ in Eq.(4) are 4, 2, 0.01 respectively. The numbers of positive templates and negative templates are 20 and 80 respectively. The learned dictionary includes 50 items.

5.2 Quantitative comparison

For quantitative performance comparison, two popular evaluation criteria are used, namely, center location error (CLE) and tracking success rate (TSR). The CLE is computed as the distance between the predicted center position and the ground truth center position. Table.1 summarizes the average center location errors in pixels.

Table 1. Average center location error (in pixel). The best two results are shown in red, blue fonts

	deer	car11	gir	cliff	skating	singer1	caviar	football	occlu2
ivt	127.5	2.106	48.47	24.81	11.72	8.483	65.96	13.61	10.21
L1	171.4	33.25	62.43	49.60	163.7	4.570	65.67	18.17	11.12
pn	25.65	25.11	23.15	11.25	55.68	32.69	44.45	13.54	18.59
vtd	11.92	27.05	21.44	34.56	13.32	4.057	58.20	4.300	10.41
mil	66.46	43.47	32.21	13.35	161.7	15.17	100.2	13.66	14.06
mtt	15.86	2.802	23.89	46.17	195.6	16.62	64.99	9.842	8.65
scm	10.02	2.520	**10.02**	**7.706**	15.15	**3.211**	2.898	**3.899**	4.44
ct	19.85	8.352	32.93	23.42	186.5	13.26	35.79	8.138	22.17
lsst	10.05	1.870	73.11	23.31	120.0	3.506	3.073	7.574	**3.70**
onnd	**8.443**	**1.742**	27.88	29.61	**7.303**	12.34	63.34	20.37	4.26
Our	**8.601**	**1.581**	**10.68**	**2.612**	11.81	**3.127**	2.486	**3.87**	**3.942**

The TSR is computed as the ratio of the number of frames the target is successfully tracked to the total frames in the sequence. To define whether the target is successfully tracked at a frame, we use the score in the PASCAL VOC challenge [23], which can be computed as

$$score = {area(R_T \cap R_G)}\big/{area(R_T \cup R_G)} \quad (13).$$

where R_T is the current the tracking result and R_G is the ground truth. Table.2 and Figure.2 give the average tracking success rates and relative tracking success rates respectively. Overall, the proposed tracker performs well against the other state-of-the-art algorithms.

Table 2. Average tracking success rate. The best two results are shown in red, blue fonts

5.3 Qualitative comparison

	deer	car11	gir	cliff	skating	singer1	caviar	football	occlu2
ivt	0.27	0.81	0.42	0.56	**0.68**	0.66	0.14	0.55	0.59
L1	0.04	0.44	0.33	0.19	0.09	0.70	0.13	0.57	0.67
pn	0.41	0.38	0.58	0.38	0.12	0.41	0.16	0.50	0.49
vtd	0.58	0.43	0.51	0.33	0.57	0.79	0.15	0.61	0.59
mil	0.21	0.17	0.52	0.46	0.12	0.34	0.13	0.57	0.61
mtt	0.52	0.81	0.63	0.31	0.09	0.42	0.14	0.66	0.72
scm	0.61	0.69	**0.68**	**0.65**	0.54	**0.86**	**0.83**	**0.83**	**0.83**
ct	0.53	0.53	0.51	0.39	0.05	0.34	0.17	**0.69**	0.54
lsst	0.58	0.81	0.12	0.56	0.13	0.79	**0.85**	0.68	0.68
onnd	**0.61**	**0.82**	0.42	0.35	**0.63**	0.20	0.05	0.41	0.79
Our	**0.62**	**0.84**	**0.69**	**0.79**	0.57	**0.87**	**0.83**	**0.83**	**0.82**

There are blurred images in deer sequence, which is difficult for most trackers to solve this situation. From Figure 1(a), we can see that when the head of the fawn becomes blurred at the frame 25 or 42, most tracking algorithms fail to follow the target, such as MIL, PN etc. The proposed algorithm successfully tracks the target object throughout the sequence. Its located accuracy and overlap rate are better than SCM, LSST, less than ONND.

In the car11 sequence, a car is driven into a very dark environment. The contrast between the tracked target and its surrounding background is low, and the ambient light changes significantly. The tracking results are illustrated in Figure 1(b). IVT, LSST, MTT and ONND algorithms perform well as our tracker in the whole sequence. However, the accuracy and robustness of these methods are less than the proposed tracker.

The tracking object in the girl sequence undergoes occlusion (complete occlusion of the girl's face as she swivels in the chair), large pose change, and scale variation with in-plane and out-of-plane rotations (from large to small, and from small to large). The tracking results are shown in Fig. 1(c). The experimental results demonstrate that our method achieves the best performance in this sequence..

In the cliffbar video, the background has similar texture to the target. Moreover, the target undergoes scale variance, in-plane rotation, and abrupt motion as shown in Figure 1(d). The L1, IVT, CT, MIL, LSST, ONND, SCM methods drift to the cluttered background, while our proposed tracker has the best performance on this sequence, it can adapt the scale and rotation change of the target, and overcome the influence of similar background and motion blur.

In the *skating* sequence, there are abrupt object motion, occlusions, severe illumination and scale changes, viewpoint changes, which lead most of the trackers to fail, especially at frame 310. Only VTD, ONND and our proposed trackers can handle these changes well and track the target throughout the sequence, as shown in Figure 1(e).

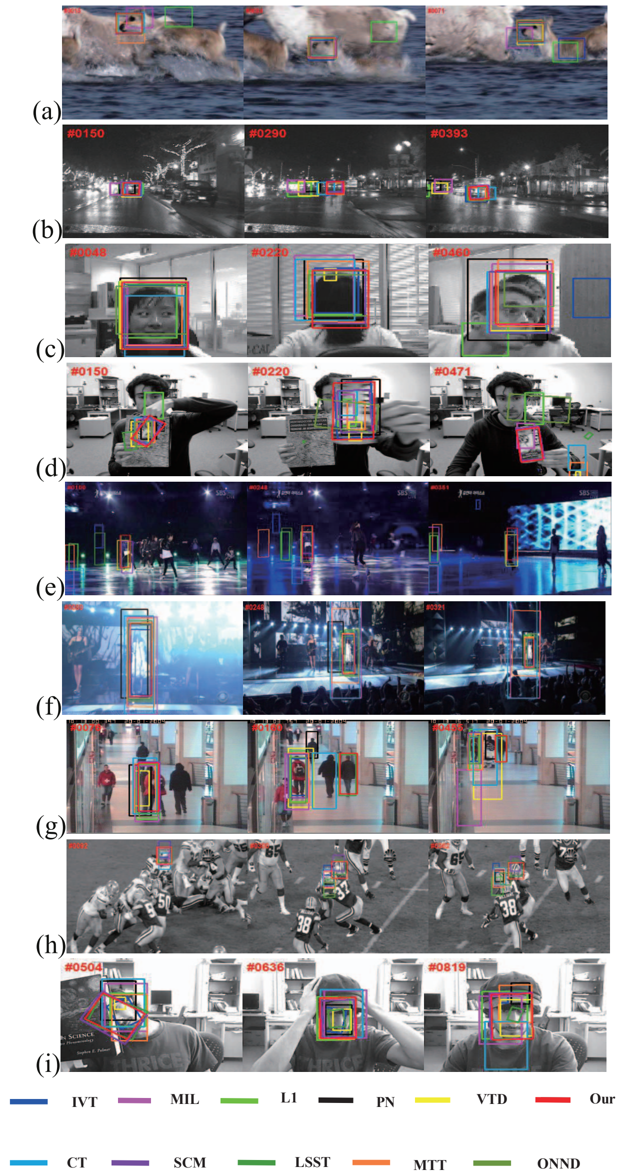

Figure 1 Comparison of 10 trackers on 9 video sequences in terms of bounding box reported

The singer1 sequence contains abrupt object motion with significant illumination and scale changes, especially, from frame 121 and frame 321, the stage light changes drastically, which cause some of the trackers to drift as shown in Fig. 1(f). SCM algorithm performs well as our tracker in the whole sequence. However, the center error and overlap rate in Table 1 and 2 have verified that our proposed trackers are better than other methods.

In the caviar sequence, the target is occluded by two people at times and one of them is similar in color and shape to the target. Numerous methods fail to track the target because there are similar objects around it when heavy occlusion occurs (such as frame 90, 120,442). In contrast, our tracker achieves stable performance in the entire sequence when there is a large scale change with heavy occlusion at frame 442.

The football sequence is challenging due to the cluttered background, because there are many football players with the similar helmets in this scene. When the tracked target approaches other football players, some trackers are not robust and begin to drift. Especially, when the two football players collide at frame 290, most tracking methods cannot locate the target correctly. Only our tracker, CT, VTD, and ONND overcome this problem and successfully locate the correct object in the whole sequence.

For the *faceocc2* sequence in Figure 1(i), most trackers start drifting from the man's face when it is almost fully occluded by the book. The proposed algorithm performs well especially when partial occlusion or in-plane rotation occurs.

6. CONCLUSIONS

In this paper, combined with the multi task sparse learning, we present a supervised approach to online learn and update a structured sparse and discriminative representation for object tracking. This approach exploits the strength of label information and encourages images from the same class to have similar representations. Experimental results on challenging image sequences demonstrate that the proposed tracker performs favorably against some state-of the- art algorithms. Possible future work is to develop the structured and low-rank representation for robust object tracking.

Acknowledgment

This work has been funded by the Natural Science Foundation of China (no. 61203270) and 2012 StateKey Laboratory of Robotics Open Project.

7. REFERENCES

[1] A. Yilmaz, O. Javed, M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol.38, no.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*, vol.46, no.7, pp. 1772-1788, 2012.

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

[6] J. Kwon and K. Lee, "Visual tracking decomposition," In *CVPR*, pp. 1269–1276, 2010.

[7] X. Mei and H. Ling, "Robust visual tracking using l1 minimization," In *ICCV*, pp. 1436–1443, 2009.

[8] D. Wang, H. Lu, and M.-H. Yang, "Least Soft-threshold Squares Tracking," In *CVPR*, pp. 2371-2378, 2013.

[9] C. Bao, Y.Wu, H. Ling, and H. Ji. Real time robust $\ell 1$ tracker using accelerated proximal gradient approach. In *CVPR*, pp.1830–1837, 2012.

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

[11] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," In *CVPR*, pp. 1838–1845, 2012.

[12] T. Zhang, B. Ghanem, S. Liu, & N. Ahuja, "Low-rank sparse learning for robust visual tracking," In *ECCV*, pp. 1–8, 2012.

[13] T. Zhang, B. Ghanem, S. Liu, & N. Ahuja, "Robust visual tracking via multi-task sparse learning," In *CVPR*, pp. 1-8, 2012.

[14] N. Wang, J. Wang, D. Yeung, "Online Robust Non-negative Dictionary Learning for Visual Tracking," In *ICCV*, 2013.

[15] 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.

[16] Z. Kalal, J. Matas, K. Mikolajczyk, "pn learning: boot-strapping binary classifiers by structural constraints," In *CVPR*, pp. 49–56, 2010.

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

[18] K. Zhang, L. Zhang, M. H. Yang, "Real-Time Compressive Tracking," In *ECCV,* pp. 864-877, 2012.

[19] E. Candes and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406–5425, 2006.

[20] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing,* vol. 54, no.1, pp. 4311-4322, 2006.

[21] Z. Jiang, Z. Lin, Larry S. Davis, "Label Consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.11, pp. 2651-2664, 2013.

[22] J. Marial, F. Bach, J. Ponce, G. Sapiro. Online dictionary learning for sparse coding, In *ICML*, 2009.

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