2016 IEEE 22nd International Conference on Parallel and Distributed Systems

Online Object Tracking Based on Convex Hull Representation

Chunjuan Bo[a,b], Dong Wang[a]
[a]School of Information and Communication Engineering, Dalian University of Technology, Dalian, China
[b]College of Electromechanical Engineering, Dalian Nationalities University, Dalian, China
(*bcj@dlnu.edu.cn, wdice@dlut.edu.cn*)

***Abstract*— This paper presents a novel tracking algorithm based on the convex hull representation model with sparse representation. The tracked object is assumed to be within the object convex hull and the candidate convex hull in the meanwhile. The object convex hull consists of a principle component analysis (PCA) subspace, and the candidate convex hull is constructed by all candidate samples with the sparsity constraint. Then we propose the objective function for our convex hull representation model, and design an iterative algorithm to solve it effectively. Finally, we present a tracking framework based on the proposed convex hull model and a simple online update scheme. Both qualitative and quantitative evaluations on some challenging video clips show that our tracker achieves better performance than other state-of-the-art methods.**

***Keywords*-object tracking; image set; convex hull; sparse representation**

I. Introduction

In general, an online object tracking system consists of two fundamental components: motion model and observation model. The motion model aims to describe the states of the tracked object over time and generates a series of possible candidate states. The Kalman filter and particle filter are common used as motion models in many classical or recent tracking algorithms (e.g., [1], [2]). The observation model usually plays the most key role of any tracking algorithm, which focuses on representing the tracked object and evaluating the likelihood of each candidate state in the current frame.

From the perspective of observation model, existing tracking algorithms can be categorized into methods based on templates (e.g., [3], [4], subspace algorithms (e.g., [5], [6], [7]), sparse representation models (e.g., [8], [9], [10]), online classifiers (e.g., [11], [12], [13]) and so on. Recently, the trackers based on linear representation models (linear subspaces or linear sparse representation models) have drawn increasing attentions and achieved promising performance. Ross *et al.* [5] develops an incremental visual tracking (IVT) method by using an incremental PCA model to capture the appearance changes of the tracked object. The experimental evaluations show that the IVT method is robust to scale change, illumination variation and pose change, but performs poorly when the target suffers from partial occlusion and background clutter during the tracking process. Inspired by the recent breakthrough of sparse representation and compressive sensing [14], Mei *et al.* [8] proposes a novel L1 tracker based on sparse representation. The L1 tracker adopts a set of object and trivial templates to sparsely represent the tracked object, where object templates describe the appearance of the target, trivial templates depict potential outliers (e.g., occlusions) and the L1 regularization imposes the sparsity constraint for solving the representation model.

Motivated by sparse representation and subspace model, in this paper we propose a novel tracking method based on the convex hull representation model with sparse representation, the contribution of which are three folds. First, we introduce the set-based representation model [15] into the tracking field, and exploit two convex hulls to depict object templates and candidates representatively. In addition, we propose a convex optimization problem for the proposed convex hull representation model, and design an iterative algorithm to solve it effectively. Finally, we develop a tracking framework by using the proposed convex hull representation model and a simple online update manner. Experiments on challenging image sequences with comparisons to state-of-the-art tracking methods show the effectiveness of the proposed tracking method.

II. The Particle Filter framework

The particle filter technique is very common used for solving the tracking problem (e.g., [5], [8], [16]) as online object tracking is a typical dynamic state inference problem. For the tracking problem, we use $\mathbf{x}_t$ to denote the state variable describing the affine motion parameters of the tracked object and adopt $\mathbf{y}_t$ to stand for its corresponding observation vector (the subscript t is the frame index). The prediction and update steps recursively estimate the posterior probability based on the following two rules:

$$p\left(\mathbf{x}_t|\mathbf{Y}_{t-1}\right) = \int p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1}\right) d\mathbf{x}_{t-1}, \quad (1)$$

$$p\left(\mathbf{x}_t|\mathbf{Y}_t\right) = p\left(\mathbf{y}_t|\mathbf{x}_t\right) p\left(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1}\right) / p\left(\mathbf{y}_t|\mathbf{Y}_{t-1}\right), \quad (2)$$

where $\mathbf{X}_t = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t]$ stands for all available state vectors up to time t and $\mathbf{Y}_t = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t]$ denotes their corresponding observations. In the tracking problem, $p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)$ is usually called the motion model that describes the state transition between consecutive frames, and

DOI 10.1109/ICPADS.2016.162

$p(\mathbf{y}_t|\mathbf{x}_t)$ denotes the observation model which evaluates the likelihood of an observed image patch belonging to the object class.

By using the particle filter technique, the posterior $p(\mathbf{x}_t|\mathbf{Y}_t)$ can be approximated by N weighted particles, which are drawn from an importance distribution $\left\{\mathbf{x}_t^i, w_t^i\right\}_{i=1,2,...,N}$, and weights of particles are updated as

$$w_t^i = w_{t-1}^i \left[p\left(\mathbf{y}_t|\mathbf{x}_t^i\right) p\left(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i\right) / q\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \mathbf{Y}_t\right)\right] \quad (3)$$

Motivated by [5], this work adopts $q(\mathbf{x}_t|\mathbf{X}_{t-1}, \mathbf{Y}_t) = p\left(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i\right)$, which is assumed to follow a Gaussian distribution. To be specific, six parameters of the affine transform are used (i.e., $\mathbf{x}_t = \{x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t\}$, where $x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t$ denote x, y translations, rotation angle, scale, aspect ratio, and skew respectively). The state transition is formulated by random walk, i.e., $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1}, \psi)$, where $\psi = diag\left(\sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_s^2, \sigma_\alpha^2, \sigma_\phi^2\right)$ is a diagonal covariance matrix. Finally, the state $\mathbf{x}_t$ is estimated as $\mathbf{x}_t^* = \sum_{i=1}^{N} w_t^i \mathbf{x}_t^i$.

III. Online object tracking based on the convex hull model with sparse representation

A. Convex hull representation

The convex hull representation model is firstly proposed in image set classification for face recognition [15], in which the query sample is supplied as a set of images rather than a single image. Let $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m] \in \mathbb{R}^{d\times m}$ denote an image set that consists of M image vectors (also called points), the convex hull representation model is defined as a linear combination of its points (equation (4)).

$$H_{\mathbf{Z}}^{\boldsymbol{\alpha}} = \left\{\mathbf{z} = \sum_{i=1}^{m} \alpha_i \mathbf{z}_i\right\} \quad (4)$$

This representation model can be also expressed as

$$H_{\mathbf{U},\mu}^{\boldsymbol{\beta}} = \left\{\mathbf{z} = \boldsymbol{\mu} + \sum_{i=1}^{k} \beta_i \mathbf{u}_i\right\}, \quad (5)$$

where $\boldsymbol{\mu}$ stands for the average vector of all data points $\boldsymbol{\mu} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{z}_i$, $\mathbf{U}$ is the orthonormal basis vectors obtained from the singular value decomposition (SVD) of the centered data matrix $[\mathbf{z}_1 - \boldsymbol{\mu}, \mathbf{z}_2 - \boldsymbol{\mu}, ..., \mathbf{z}_m - \boldsymbol{\mu}]$, k is the number of eigenvectors ($k \ll m$).

B. The proposed convex hull representation model for online object tracking

The Figure 1 explains the basic ideas of the proposed method, from which we can see that there exist two convex hull models for the tracking problem. The first convex hull describes the object model, which assume that the tracked object is within the convex hull $H_{\boldsymbol{\mu},\mathbf{U}}^{\boldsymbol{\alpha}}$ (i.e., the tracked object $\mathbf{y}$ can be represented by a PCA subspace centered by $\boldsymbol{\mu}$ and

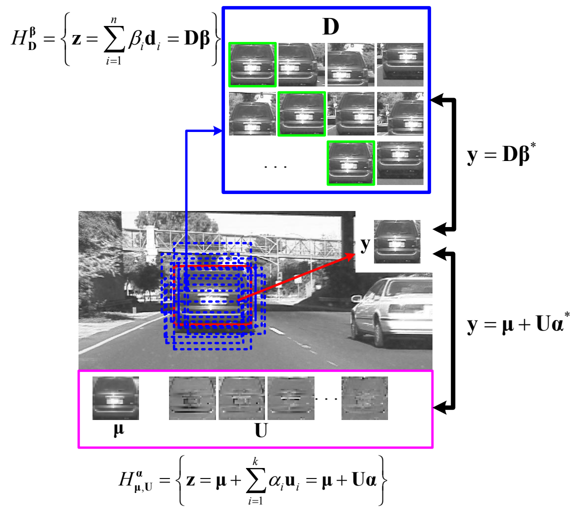

Figure 1. An illustration of our convex hull representation model for visual tracking.

spanned by $\mathbf{U}$, $\mathbf{y} = \boldsymbol{\mu} + \mathbf{U}\alpha^*$). The second convex hull is the candidate dictionary in the current frame, which assume that the tracked object is within the convex hull $H_{\mathbf{D}}^{\boldsymbol{\beta}}$ (i.e., the tracked object can be linearly represented by candidates samples in a given frame, $\mathbf{y} = \mathbf{D}\beta^*, \beta^* \succeq \mathbf{0}$). Here we adopt non-negative constraints on the coding coefficients to make sure that each coefficient can be explained as the similarity between the tracked object and its corresponding candidate. We note that the coefficient vector $\boldsymbol{\beta}$ should be sparse as some candidates are related to the tracked object (see the green bounding boxes in Figure 1); the remaining ones are related to background samples. Thus, in this work, we use the L1-norm regularization to encourage the sparsity of $\boldsymbol{\beta}$ and propose the following optimization problem.

$$\begin{aligned}&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \tfrac{1}{2}\left\|\boldsymbol{\mu} + \mathbf{U}\boldsymbol{\alpha} - \mathbf{D}\boldsymbol{\beta}\right\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \\ &s.t.\boldsymbol{\beta} \succeq \mathbf{0}\end{aligned} \quad (6)$$

In order to handle outliers (e.g., partial occlusion), we introduce a mask matrix $\mathbf{W}$ to remove the outliers from the proposed convex hull representation model (how to obtain the matrix $\mathbf{W}$ is presented in Section 3.3). Thus, the optimization problem (6) can be converted as

$$\begin{aligned}&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \tfrac{1}{2}\left\|\mathbf{W} \odot (\mu + \mathbf{U}\boldsymbol{\alpha} - \mathbf{D}\boldsymbol{\beta})\right\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \\ &s.t.\boldsymbol{\beta} \succeq \mathbf{0}\end{aligned} \quad (7)$$

where $\odot$ denotes the element-wise multiplication operator.

Let $\boldsymbol{\mu}' = \mathbf{W} \odot \boldsymbol{\mu}$, $\mathbf{U}' = \mathbf{W} \odot \mathbf{U}$ and $\mathbf{D}' = \mathbf{W} \odot \mathbf{D}$, the optimization problem (7) can be rewritten as

$$\begin{aligned}&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} J(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &s.t.\boldsymbol{\beta} \succeq \mathbf{0}\end{aligned}, \quad (8)$$

where

$$J(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{2}\left\|\boldsymbol{\mu}' + \mathbf{U}'\boldsymbol{\alpha} - \mathbf{D}'\boldsymbol{\beta}\right\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 = \frac{1}{2}\left\|\boldsymbol{\mu} - [-\mathbf{U}', \mathbf{D}']\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix}\right\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \tag{9}$$

It can be seen from equation (9), the objective function $J(\boldsymbol{\alpha},\boldsymbol{\beta})$ consists of two components: a standard least squares criterion on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and an L1 regularization term on $\boldsymbol{\beta}$. In order to solve it effectively, we design an iterative algorithm based on the following two propositions.

Proposition 1 Given the optimal coefficient vector $\boldsymbol{\alpha}^*$, the optimal coefficient vector $\boldsymbol{\beta}^*$ can be solved by a standard L1-norm regularized sparse coding algorithm.

If $\boldsymbol{\alpha}^*$ is known, the optimization problem of equation (8) is equivalent to the minimization of

$$F(\boldsymbol{\beta}) = J(\boldsymbol{\alpha}^*,\boldsymbol{\beta}) = \frac{1}{2}\left\|(\boldsymbol{\mu}' + \mathbf{U}'\boldsymbol{\alpha}^*) - \mathbf{D}'\boldsymbol{\beta}\right\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1, \tag{10}$$

with the non-negative constraint $\boldsymbol{\beta} \succeq \mathbf{0}$. This is a standard L1-regularization problem, which can be solved by the LASSO method effectively.

Proposition 2 Given the optimal coefficient vector $\boldsymbol{\beta}^*$, the optimal coefficient vector $\boldsymbol{\alpha}^*$ can be solved by a standard least squares regression method.

If $\boldsymbol{\beta}^*$ is known, the optimization problem of equation (8) is equivalent to the minimization of

$$G(\boldsymbol{\alpha}) = \frac{1}{2}\left\|(\mathbf{D}'\boldsymbol{\beta}^* - \boldsymbol{\mu}') - \mathbf{U}'\boldsymbol{\alpha}\right\|_2^2 = J(\boldsymbol{\alpha},\boldsymbol{\beta}^*) - \lambda\|\boldsymbol{\beta}^*\|_1. \tag{11}$$

This is a simple least squares problem, and the solution can be easily found obtained by setting the derivative $\partial G(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$ to zero,

$$\partial G(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha} = (\mathbf{U}')^\top\left[(\mathbf{D}'\boldsymbol{\beta}^* - \boldsymbol{\mu}') - \mathbf{U}'\boldsymbol{\alpha}\right] = \mathbf{0}. \tag{12}$$

Then $\boldsymbol{\alpha}^* = \left[(\mathbf{U}')^\top\mathbf{U}'\right]^{-1}(\mathbf{U}')^\top(\mathbf{D}'\boldsymbol{\beta}^* - \boldsymbol{\mu}')$ is the analytic solution of the optimization problem (11).

By **Propositions 1** and **2**, the optimization problem (8) can be solved iteratively.

C. Other details

Based on the proposed convex hull representation model, we design a generative tracking algorithm. Here we introduce how to obtain the mask $\mathbf{W}$ in detail (we drop the frame index t for clarity). The mask $\mathbf{W}$ is a diagonal matrix $\mathbf{W} = diag(w_1, w_2, ..., w_d)$, and w_i is a binary value for every i. $w_i = 1$ indicates that the i-th pixel is an inlier; and $w_i = 0$ indicates that the i-th pixel is an outlier.

$$w_i = \begin{cases} 1, & |y_i^o - \mu_i| < \theta \\ 0, & otherwise \end{cases}, \tag{13}$$

where $\mathbf{y}^o = [y_1^o; y_2^o; ...; y_d^o]$ denotes the observation image vector, $\boldsymbol{\mu} = [\mu_1; \mu_2; ...; \mu_d]$ stands for the average vector of the PCA model, and θ is a per-defined threshold to identify inliers and outliers.

Then we recovery the observation vector by replacing the outliers with its corresponding parts of the average vector,

$$\mathbf{y}^r = \mathbf{W}\mathbf{y}^o + (\mathbf{I} - \mathbf{W})\boldsymbol{\mu}, \tag{14}$$

where $\mathbf{I}$ denotes an identity matrix. Finally, the recovered observation patch is cumulated and then used to update the PCA model by using the incremental PCA method [5].

IV. Experiments

A. Implementation details

Each observed image patch is resized to 32×32 pixels. As a tradeoff between effectiveness and efficiency, 16 eigenvectors are adopted to form the PCA subspace and 600 particles are sampled to build the candidate dictionary. The regularization parameter λ is set as 0.05, and the threshold parameter θ is chosen as 0.08.

B. Qualitative and Quantitative comparisons

Figure 2 demonstrates some selected tracking results. Table II summarizes the average center location error of these trackers, where a smaller error indicates a better performance. From Table II, we can see that our tracker performs well against the other state-of-the-art tracking algorithms on many challenging video clips.

V. Conclusions

In this work, we present a novel online object tracking method based on the convex hull representation model with sparse representation. First, we adopt two convex hulls to depict the object templates and candidate samples, and point out their relationship for the tracking problem. Second, we propose a convex optimization problem to exploit the proposed convex hull representation model for visual tracking, and design an iterative algorithm to solve this problem effectively. Third, based on our convex hull representation model, we develop a tracking framework by combining it with an additional model update manner. The experimental results show that our tracker achieves better performance than other competing tracking methods.

Acknowledgement This work was supported in part by the Natural Science Foundation of China under Grants No. 61502070, in part by Fundamental Research Funds for Central Universities under Grant No. DUT16RC(4)16, and in part by Fundamental Research Funds for Central Universities under Grant No. DCPY2016107.

References

[1] D. Comaniciu, V. R. Member, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–575, 2003.

[2] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *ECCV*, 2002, pp. 661–675.

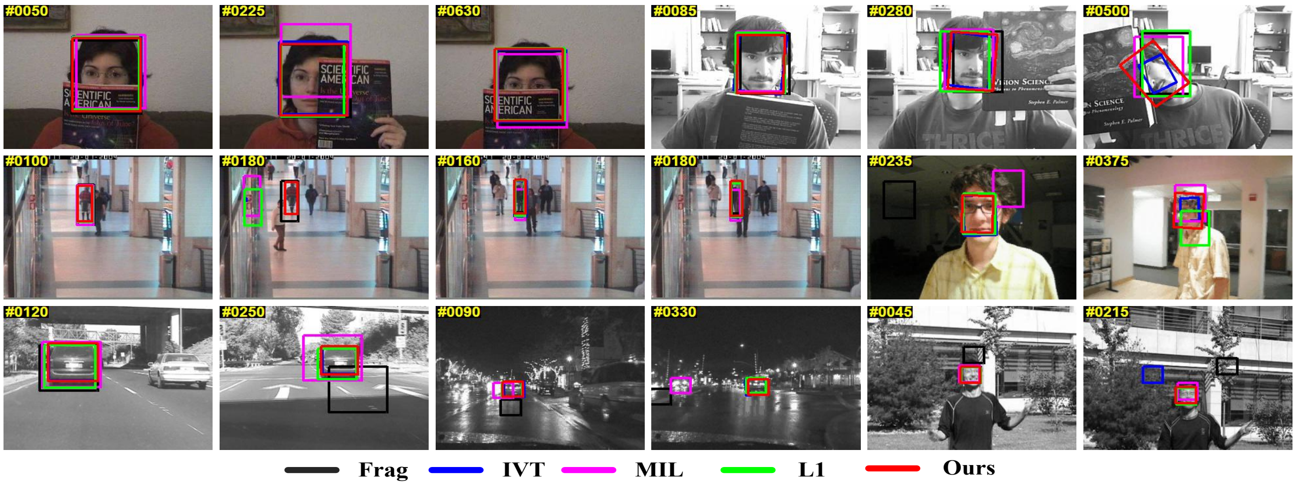

Figure 2. The representative results of different trackers.

Table I
AVERAGE CENTER LOCATION ERRORS OF TRACKING ALGORITHMS. THE BEST TWO RESULTS ARE SHOWN IN RED AND BLUE FONTS.

Sequence	FragT [3]	IVT [5]	MIL [11]	VTD [4]	TLD [12]	L1 [8]	APGL1 [17]	MTT [18]	Ours
FaceOcc1	5.6	9.2	32.3	11.1	17.6	6.5	6.8	14.1	6.1
FaceOcc2	15.5	10.2	14.1	10.4	18.6	40.2	6.3	19.2	4.5
Caviar1	5.7	45.2	48.5	3.9	5.6	119.9	50.1	20.9	1.5
Caviar2	5.6	8.6	70.3	4.7	8.5	3.2	63.1	65.4	1.7
DavidIndoor	148.7	3.1	34.3	49.4	13.4	25.6	10.8	13.4	2.8
Car1	179.8	2.9	60.1	12.3	18.8	4.1	16.4	37.2	2.7
Car2	63.9	2.1	43.5	27.1	25.1	33.3	1.7	1.8	2.6
Jumping	58.4	36.8	9.9	63.0	3.6	92.4	8.8	19.2	6.1
Average	60.4	14.8	39.1	22.7	13.9	40.7	20.5	23.9	3.5

[3] A. Adam, E. Rivlin, and I. Shimshoni, “Robust fragments-based tracking using the integral histogram,” in *CVPR*, 2006, pp. 798–805.

[4] J. Kwon and K. M. Lee, “Visual tracking decomposition,” in *CVPR*, 2010, pp. 1269–1276.

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

[6] D. Wang, H. Lu, and M.-H. Yang, “Least soft-threshold squares tracking,” in *CVPR*, 2013, pp. 2371–2378.

[7] X. Li, W. Hu, Z. Zhang, X. Zhang, M. Zhu, and J. Cheng, “Visual tracking via incremental log-Euclidean Riemannian subspace learning,” in *CVPR*, 2008, pp. 1–8.

[8] X. Mei and H. Ling, “Robust visual tracking using ℓ_1 minimization,” in *ICCV*, 2009, pp. 1436–1443.

[9] X. Jia, H. Lu, and M.-H. Yang, “Visual tracking via adaptive structural local sparse appearance model,” in *CVPR*, 2012, pp. 1822–1829.

[10] W. Zhong, H. Lu, and M.-H. Yang, “Robust object tracking via sparsity-based collaborative model,” in *CVPR*, 2012, pp. 1838–1845.

[11] B. Babenko, M.-H. Yang, and S. Belongie, “Visual tracking with online multiple instance learning,” in *CVPR*, 2009, pp. 983–990.

[12] Z. Kalal, K. Mikolajczyk, and J. Matas, “Tracking-learning-detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.

[13] K. Zhang, L. Zhang, and M.-H. Yang, “Real-time compressive tracking,” in *ECCV*, 2012, pp. 864–877.

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[15] H. Cevikalp and B. Triggs, “Face recognition based on image sets,” in *CVPR*, 2010, pp. 2567–2573.

[16] D. Wang, H. Lu, and M.-H. Yang, “Online object tracking with sparse prototypes,” *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 314–325, 2013.

[17] C. Bao, Y. Wu, H. Ling, and H. Ji, “Real time robust ℓ_1 tracker using accelerated proximal gradient approach,” in *CVPR*, 2012, pp. 1830–1837.

[18] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, “Robust visual tracking via multi-task sparse learning,” in *CVPR*, 2012, pp. 2042–2049.