ROBUST VISUAL TRACKING VIA ADAPTIVE STRUCTURE-ENHANCED PARTICLE FILTER

Nan Song[1], *Kezhi Li*[2], *Wei Chen*[1,3]
[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China
[2]Dept. of Electrical and Electronic Engineering, Imperial College London, United Kingdom
[3]Beijing Engineering Research Center of High-speed Railway Broadband Mobile Communications
Corresponding author: *Wei Chen*
15120129@bjtu.edu.cn, kezhi.li@imperial.ac.uk, weich@bjtu.edu.cn

ABSTRACT

An effective representation model plays an important role in the visual tracking, as it relates to how the most meaningful information are recognized and understood in the dictionary space. However, it is difficult to know the structure and the weights of tracking objects in advance. In addition, how to balance the adaption and robustness in tracking algorithms remains a nontrivial problem. In this paper, we propose a robust visual tracker based on adaptive structure-enhanced regularizations, and achieve a sequential Monte Carlo searching via simplified particle filters. Specifically, multiple atomic norms are incorporated in the cost function in the target dictionary space, and their weights are updated adaptively during the detection step between each frame. Sparse and low-rank structures as well as other atomic norms enhance the robustness by capturing various features meanwhile ruling out outliers, and the velocity of moving objects are considered accordingly in the probabilistic distribution of particles. Moreover, the algorithm has been accelerated by adopting prefilters as classifiers for target particles using pixel variances in colours and intensities, which ensures a real-time tracking in practice. On challenging tracking datasets, the proposed approach show advantages in tracking fast-moving objects and favorable performance against other 10 state-of-the-art visual trackers.

***Index Terms*—** Visual tracking, atomic norm representation, dictionary learning, particle filter.

I. INTRODUCTION

Visual tracking is an attractive technology owing to its wide applications including vehicle tracking [1], surveillance [2], medical diagnosis [3] and video information compression [4]. Detection algorithm and sparse signal processing [5]–[7] have also been greatly developed. The accuracy of the parameters directly affects the performance of target detection, and the acquisition of parameters is usually achieved by feature matching algorithm. Commonly features are texture feature, SIFT feature, SURF feature, Harris-Laplacian. Zdenek Kalal *et al.* [8] propose a detection algorithm based on tracking-learning-detection (TLD) with the detection algorithm to solve the problem of the tracked object shape change and partial occlusion. Oron *et al.* [9] use the TLD framework for pedestrian target tracking.

However, there still exist challenges in practical applications [10]–[12]. It is difficult to obtain the tracking target structures in advance, and the performance of tracking can be significantly affected by lightning changes, morphological changes, background noise and other factors. These conditions are different in various scenarios. In addition, different applications have distinct requirements in robustness, accuracy and computational complexity. In this paper, we propose a novel approach for visual tracking. Firstly, we extend the regularizations in the estimation equation to multiple atomic norms, which imposes an adaptive penalties on the dictionary spaces and enhances the capability of the tracker in recognizing targets with unknown structures. Secondly, the weights of each penalty terms in the detection optimization are updated between each adjacent frame, in which a gradual changing target can be comprehensively considered. Furthermore, a simplified particle filter is implemented to capture the target probabilistic distributions by building particles classified in pixel colours, intensities, and target velocities. It strengthens the approach to the real-time efficiency. Finally, the tracking accuracy of our algorithm is verified by experiments on a large number of visual datasets. The results come out in front in the test.

The remainder of the paper is organized as follows. Section II shows the process of establishing a mathematical model for tracking. Section III provides the experimental of the paper. Conclusions are drawn in Section IV.

II. RELATED WORK

Roughly speaking, tracking algorithms can be divided into two categories: discriminative methods and generative methods. Discriminative methods treat visual tracking as a classification problem of foreground and background. The generative methods adopts online updating method to adjust the parameters so as to meet the change of the target appearance. IVT [13] propose view the model to adapt to the change in the target as the goal of learning online. Kwon *et al.* In [14], the improved particle filter, multiple

This work is supported by the Natural Science Foundation of China (61671046, 61725101); the Beijing Natural Science Foundation (4182050); the State Key Laboratory of Rail Traffic Control and Safety (RC-S2018ZZ002) of Beijing Jiaotong University; UK Medical Research Council through grant MC-A658-5TY30.

observations and motion modes are combined to handle large appearance and movement changes. For the first time, the theory of sparse representation is applied to the target tracking and Mei *et al.* [15] propose the l_1 tracker. But it needs long time to solve l_1 norm optimization problem. Zhang *et al.* [16] use the low rank representation, and propose a target tracking algorithm based on low rank sparse learning (CLRST). In Zhang's model [16], the target template is composed of gray scale pixel vectors. The defection of the target template does not contain rich image information, and there are correlations between the target templates, which is not conducive to sparse representation.

III. DETECTION AND TRACKING

Firstly, the target detection algorithm is used to find the moving target, and features in the dictionary space are updated accordingly. When the tracked target is found, the tracking algorithm starts to work, and the dictionary and weights are changing adaptively. After that, the detection algorithm helps tracking the target in each frame.

III-A. Detection

The detection uses moving window for detecting objects to see whether the target appears in each frame of the video, and then finds the scanning window containing the target. In each frame, the algorithm produces a large number of moving windows as particles according to the calculated position distribution (see Sec. III-B). We select the range of windows based on the tracking results of the previous frame considering the target velocity and see if window contain the target. After that, we implement simple classifiers to filter the windows aiming at improving the detection efficiency. Two classifiers are adopted in the algorithm, variance classifier and colour classifier. Variance classifier uses a threshold of pixel's variance in the sliding window to determine whether the window contains the target. The following equation computes the variance

$$\mathbf{D}(\mathbf{p}) = \mathbf{E}(\mathbf{p})^2 - \mathbf{E}(\mathbf{p}^2), \tag{1}$$

where $\mathbf{p} \in \mathbb{R}^{m_1 \times m_2}$ denotes the pixel intensities in the considering window of size m_1 by m_2, $\mathbf{D}(\mathbf{p})$ represents the variance of pixel intensities in the window, $\mathbf{E}(\mathbf{p})^2$ represents the square of the expected value of pixel intensities in a window, and $\mathbf{E}(\mathbf{p}^2)$ represents the expectation of the square value of pixel intensities in the window. The colour classifier can be calculated in the similar approach where we need to consider 3 channels (RGB or HSV). A moving window meets the requirements of both classifiers can be kept as qualified candidates for future computations.

III-B. Particle Filter

In this paper, a recursive Monte Carlo particle filter is leveraged to implement the recursive Bayesian filter. The idea is, the posterior probability density is expressed by the estimation of the random particle set with weights, and the estimation of the state is obtained. The target image set $\mathbf{Y}_t = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t\}$ is given at the *t* moment. An iterative solution is used to estimate hidden state variables $\mathbf{x}_t$,

$$p(\mathbf{x}_t|\mathbf{y}_t) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_{t-1}) d\mathbf{x}_{t-1}, \tag{2}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ represents a dynamic model between two consecutive states, $p(\mathbf{y}_t|\mathbf{x}_t)$ represents the observation model. It is used to estimate the probability of the observed $\mathbf{y}_t$ in the case of the known state $\mathbf{x}_t$. Given some observations at time t, the optimal state of the tracking target can be obtained by estimating the maximum posterior probability of the N samples at time *t*:

$$\mathbf{x}_t = \arg\min p(\mathbf{y}_t^i|\mathbf{x}_t^i) p(\mathbf{x}_t^i|\mathbf{x}_{t-1}), i = 1, 2, ..., N, \tag{3}$$

where $\mathbf{x}_t^i$ is the *i*th sample particle of state $\mathbf{x}_t$, and $\mathbf{y}_t^i$ is the image block corresponding to $\mathbf{x}_t^i$. After the maximum posterior probability is calculated in the particle filter framework, the particle with the largest probability is selected as the tracking result in a large number of particles in time t. In addition, the result in each frame is obtained by resampling of particles based on the result in the last frame. This approach can avoid particles degeneracy to a certain extent.

III-C. Atomic Norm and Adaptive Weights

Atomic norms are considered in the optimization tracking problem in the framework of particle filter, and the regularizations are set as the nonnegative combination of atoms from a set $\mathcal{A}$.

Definition 1 (Atomic Norm): [17] The atomic norm $||\cdot||_{\mathcal{A}}$ of $\mathcal{A}$ is the Minkowski function (or the gauge function) associated with conv($\mathcal{A}$) (the convex hull of $\mathcal{A}$):

$$||x||_{\mathcal{A}} = \inf\{t > 0 | x \in t \text{ conv}(\mathcal{A})\}. \tag{4}$$

If conv($\mathcal{A}$) is compact, centrally symmetric, and contains a ball of radius $\epsilon > 0$ around the origin, the gauge function is a norm. Specifically, when $\mathcal{A} \in \mathbb{C}^n$ is the set of unit norm 1-sparse elements, the atomic norm $||\cdot||_{\mathcal{A}}$ is the l_1 norm [18]. When A is the set of unit norm 1-rank matrices, the atomic norm is the nuclear norm [19]. Researchers have showed that minimizing the atomic norm subject to constraints provided exact solutions of a variety of linear inverse problems with nearly optimal bounds on the number of measurements required [17], [20].

In the tracking process, we define $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2 ..., \mathbf{x}_n]$ and $\mathbf{x}_n$ is a vector of pixels in a window. $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2 ..., \mathbf{d}_n]$ is composed of tracking results that gets from every previous frame. We define $\mathbf{X} = \mathbf{DZ}$, $\mathbf{Z}$ has a structure that can be represented by combinations of multiple atomic norms. In Zhang's paper [16], $\mathbf{Z}$ has been shown to have low rank and sparse properties in most cases. While in our case, the concept has been extended to all structures whose atomic norms are low, and Zhang's work can be seen as a special case of ours. In the process of target tracking, the observation model plays a major role, and its purpose is to calculate the likelihood probability of the observed variable. A prediction state $\mathbf{x}_t$ corresponds to an observation target image, then the n prediction states correspond to n candidate target image blocks $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$. Then we aim at solving the low rank sparse representation problem by

$$\begin{aligned}\min_{\mathbf{Z},\mathbf{E}} \ & \lambda_1 \|\mathbf{Z}\|_{\mathcal{A}_1} + \lambda_2 \|\mathbf{Z}\|_{\mathcal{A}_2} + \cdots + \lambda_k \|\mathbf{Z}\|_{\mathcal{A}_k} \\ & + \kappa_1 \|\mathbf{Z} - \mathbf{Z}_0\|_2 + \kappa_2 \|\mathbf{E}\|_1 \quad s.t. \ \mathbf{Y} = \mathbf{DZ} + \mathbf{E}\end{aligned} \tag{5}$$

where $\mathbf{Z}$ combines multiple atomic norms such as low rank and sparse properties, $\mathbf{Z_0}$ is composed of the low atomic norm representations of the result from the previous frame, and $\mathbf{E}$ represents the error caused by the noise in many different occlusions. $||\cdot||_{\mathcal{A}_1}, \cdots, ||\cdot||_{\mathcal{A}_k}$ represent k different atomic norms, k denotes the number of atomic norms in consideration, $\lambda_1 \sim \lambda_k$, $\kappa_{1,2}$, are weights of each term. These atomic norms can be the nuclear norm which is the sum of the singular values of the matrix, or l_1 norm which is the sum of absolute values, or other atomic norms, such as moments under *Fourier* bases, *Wavelets* bases, etc [17]. Before the tracking process, people can choose the types and numbers of atomic norms in Eq. (5) according to the structure of the tracking target. Normally 2 or 3 regularizations can be adopted in the Eq. (5), and atomic norms including l_1, nuclear norm, *Fourier* and *wavelets* are highly recommended if no prior information of the tracking target is provided. Because most images can be decomposed to such dictionaries naturally and efficiently.

Algorithm 1 Atomic Norm Tracking With Detection Algorithm.

Input: Dictionary template $\mathbf{D}_{t-1}$
Particles of n
Tracking result $\mathbf{y}_{t-1}$ at previous frame
Output: Tracking result $\mathbf{y}_t$
Updated dictionary $\mathbf{D}_t$
1: Detect the target using the classifier (1) according to the result $\mathbf{y}_{t-1}$ of previous frame
2: The result of the detection is used as $\mathbf{d}_t$ in the dictionary $\mathbf{D}$
3: Compute atomic norm representation of $\mathbf{Z}$ for $\mathbf{y}$ by (5)
4: The $\mathbf{Z}$ with the highest similarity score is used as the tracking result
5: Update $\lambda_1 \sim \lambda_k$, $\kappa_{1,2}$ toward the gradient direction
6: The result of the tracking is used as input to next frame

Inexact augmented lagrange multiplier (IALM) is used for low rank minimization of matrices. Its main idea is to join Lagrange in the objective function to be an augmented Lagrange function. In the iterative process it adjusts the penalty factor and the Lagrange multipliers, so as to meet the convergence condition and approach the optimal approximate solution. In order to solve Eq. (5), in the case of sparsity, we use $S_\lambda(\mathbf{Z})$ that is the soft-thresholding operator,

$$S_\lambda(Z_{ij}) = sign(Z_{ij}) \max(0, |Z_{ij}| - \lambda). \tag{6}$$

In the case of low rank we use $J_\lambda(\mathbf{Z}) = \mathbf{U} S_\lambda(\Sigma) \mathbf{V}^T$, that is the singular value soft-thresholding operator. When the atomic norm applies other norms, we use the greedy forward-backward operator to calculate it [21]. After obtaining the $\mathbf{Z}$ with the highest similarity score as the tracking result, we use an extra step to update the weights in Eq. (5). The rule of the updating is to increase the weights that can represent the target features better, and decrease the weights that fluctuate in the previous frames. We create a memory to restore the values and weights of each term in previous certain number of frames. We assume that the essential information of the dictionary $\mathbf{D}$ cannot change dramatically during the tracking process as long as it is not obscured. Based on this idea the weights can be updated adaptively, and the update speeds are chosen as less than the amount proportional to the gradient of the atomic norms.

The complete approach can be summarized in the Algorithm 1. Firstly, we can detect the moving target using variance classifier and colour classifier. Those classifiers are built from the result's pixels of the previous frame. The result of the detection is treated as a column of the dictionary $\mathbf{D}$. Secondly, we compute the atomic norm representation of $\mathbf{Z}$. The $\mathbf{Z}$ with the highest similarity score is used as the tracking result. This is the tracking stage. In the end, we update the parameters and get the results as input for the next frame.

IV. EXPERIMENTAL RESULTS

In order to verify the effectiveness of the proposed model, this section uses 10 trackers on 7 widely used sequences. Aiming at carrying out a fair comparison, we compare our algorithm to methods whose source code is available publicly. Meanwhile we try to ensure that each tracking algorithm has the similar sets of specific parameters. In recent years, benchmark datasets are used for many visual tracking. In this paper we use the same ground truth for each tracking algorithm in the datasets. For vision tracking algorithm, two evaluation methods are applied in order to evaluate the performances of the algorithms objectively and accurately. One method is the average precision (AP), the other method is the success rate (SR). The precision rate refers to the Euclidean distance between the center position (x, y) of the tracking results in each frame and the true center position (x_{true}, y_{true}). The average precision can be calculated as:

$$AP = \frac{1}{T} \sum_{t=1}^{T} \sqrt{(x_t - x_t^{true})^2 + (y_t - y_t^{true})^2}, \tag{7}$$

where (x, y) indicates the predicted center position of the result at time *t*, (x_{true}, y_{true}) indicates the ground truth position at time *t*. *T* represents the total number of video frames. The success rate refers to the ratio of the overlapping areas that are the tracking result's window and the ground truth window,

$$SR = \frac{1}{T} \sum_{t=1}^{T} \frac{w_t \cap w_t^{true}}{w_t \cup w_t^{true}}, \tag{8}$$

where w_t represents the area of the window at time *t*, w_t^{true} indicates the area of the ground truth window at time *t*.

Table I. PRECISION OF TRACKING ALGORITHM

Video	David	David2	Sylv	Fish	Shaking	Singer1	Singer2
CT [22]	0.73	0.01	0.91	0.72	0.70	0.56	0.92
IVT [13]	0.85	0.97	0.92	0.92	0.04	0.77	0.13
DFT [23]	0.37	0.72	0.90	0.82	0.74	0.62	0.59
ASLA [24]	0.37	0.80	0.91	0.67	0.19	0.62	0.85
L1APG [25]	0.72	0.96	0.91	0.41	0.09	0.35	0.04
ORIA [26]	0.55	0.86	0.92	0.56	0.52	0.84	0.09
MTT [27]	0.33	0.96	0.91	0.16	0.05	0.47	0.04
CSK [28]	0.64	0.94	0.92	0.18	0.65	0.72	0.04
TLD [8]	0.59	0.93	0.93	0.62	0.08	0.78	0.16
LRT [16]	0.78	0.95	0.91	0.80	0.85	0.71	0.55
Proposed Method	0.90	0.92	0.94	0.93	0.51	0.83	0.60

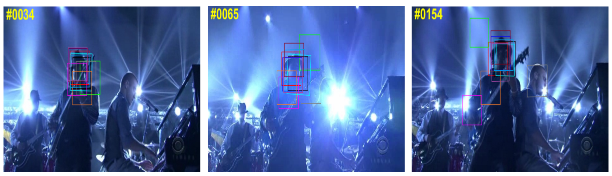


Fig. 1. Various algorithms are compared in a frame of video. The box represents the result location of the algorithms. The video is Singer1.

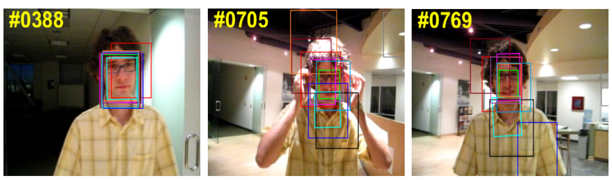
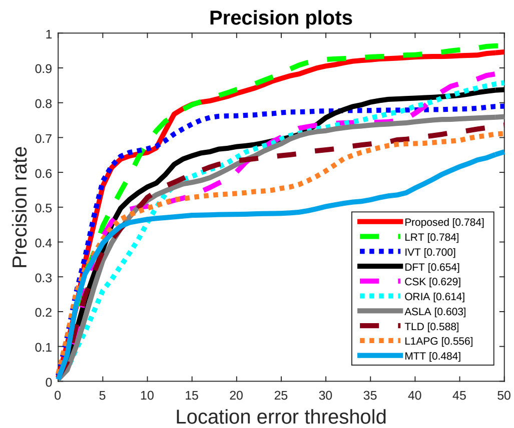


Fig. 2. Various algorithms are compared in a frame of video. The box represents the result location of the algorithms. The video is David.

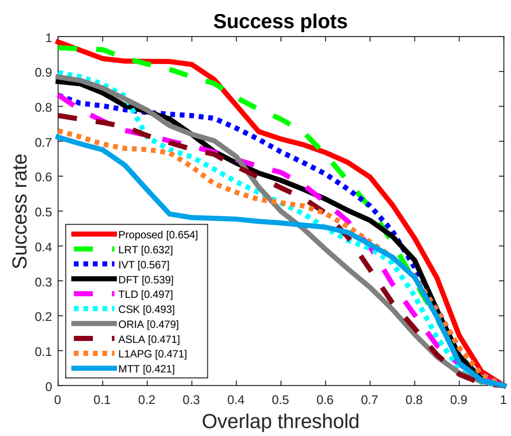
Table I shows the average precision of 11 tracking algorithms on the 7 videos. Compared with other methods, our proposed method has the best results on David, Sylv, Fish videos.

Figure 1 shows the results for the Singer1 video. This video contains scale, various illumination and changes in different perspectives. Figure 2 shows the results for the David video. The difficulty is the changes on the background. Our method has a good performance on all frames.

Figure 3(a) shows the precision curve and figure 3(b) shows the success rate curve of various algorithms on the 7 videos. The method of evaluating each algorithm is to initialize the same location and run the algorithm in each test video. Finally, we record all the results data, including precision or success rate. The proposed algorithm is the best or second best results in the figures. The precision curve is better, and the accuracy of the algorithm is higher than most of other algorithms. Similarly, the success rate curve is greater than other methods.



(a)



(b)

Fig. 3. The results of various algorithms are calculated on datasets. (*a*) Precision rates of different algorithms. (*b*) Success rates of different algorithms.

V. CONCLUSIONS

In this paper, we focus on the object detection and tracking in computer vision. Based on the particle filter framework and adaptive structure-enhanced regularizations, we propose a novel robust visual tracker, which jointly updates the target dictionary and parameters to obtain tracking results with better accuracy. Multiple atomic norms are used in the cost function in the dictionary space. The proposed approach uses the appropriate classifier to ensure the real-time tracking in practice. We evaluate the proposed approach with tracking datasets, and demonstrate the superiority of the proposed approach against other 10 state-of-the-art visual trackers.

VI. REFERENCES

[1] X. Ding, W. Chen, and I. J. Wassell, "Compressive sensing reconstruction for video: An adaptive approach based on motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 7, pp. 1406–1420, July 2017.

[2] Yi Wu, Jongwoo Lim, and Ming Hsuan Yang, "Online object tracking: a benchmark," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2411–2418.

[3] Adriaan Arie Johannes Lefeber, "Tracking control of nonlinear mechanical systems," *Universiteit Twente*, 2017.

[4] Shipra Ojha and Sachin Sakhare, "Image processing techniques for object tracking in video surveillance-a survey," in *International Conference on Pervasive Computing*, 2015, pp. 1–6.

[5] W. Chen, "Simultaneous sparse bayesian learning with partially shared supports," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1641–1645, 2017.

[6] W. Chen, D. Wipf, Y. Wang, Y. Liu, and I. J. Wassell, "Simultaneous bayesian sparse approximation with structured sparse models," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6145–6159, Dec 2016.

[7] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2016–2029, April 2013.

[8] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 34, no. 7, pp. 1409–22, 2012.

[9] Shaul Oron, Aharon Bar-Hillel, and Shai Avidan, "Real-time tracking-with-detection for coping with viewpoint change," *Machine Vision Applications*, vol. 26, no. 4, pp. 507–518, 2015.

[10] Baojie Fan, Cong Yang, Xiaomao Li, and Yandong Tang, "Structured low rank tracker with smoothed regularization," in *Visual Communications and Image Processing*, 2016, pp. 1–4.

[11] Weizhi Lu, Jinglin Zhang, Kidiyo Kpalma, and Joseph Ronsin, "Efficient visual tracking via low-complexity sparse representation," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 21, Mar. 2015.

[12] Kezhi Li and Shuang Cong, "State of the art and prospects of structured sensing matrices in compressed sensing," *Frontiers of Computer Science*, vol. 9, no. 5, pp. 665–677, Oct. 2015.

[13] David A. Ross, Jongwoo Lim, Ruei Sung Lin, and Ming Hsuan Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[14] S Oron, A Bar-Hillel, D Levi, and S Avidan, "Locally orderless tracking," pp. 1940–1947, 2012.

[15] Xue Mei and Haibin Ling, "Robust visual tracking using L1 minimization," pp. 1436–1443, 2009.

[16] Tianzhu Zhang, Si Liu, Narendra Ahuja, Ming Hsuan Yang, and Bernard Ghanem, "Robust visual tracking via consistent low-rank sparse learning," *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171–190, 2015.

[17] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec 2013.

[18] Emmanuel J. Cands, Justin K. Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[19] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[20] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Comput. Math.*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[21] N. Rao, P. Shah, S. Wright, and R. Nowak, "A greedy forward-backward algorithm for atomic norm constrained minimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 5885–5889.

[22] Kaihua Zhang, Lei Zhang, and Ming Hsuan Yang, "Real-time compressive tracking," in *European Conference on Computer Vision*, 2012, pp. 864–877.

[23] E. Learnedmiller and L. Sevillalara, "Distribution fields for tracking," in *Computer Vision and Pattern Recognition*, 2012, pp. 1910–1917.

[24] Xu Jia, Huchuan Lu, and Ming Hsuan Yang, "Visual tracking via adaptive structural local sparse appearance model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1822–1829.

[25] Hui Ji, Haibin Ling, Yi Wu, and Chenglong Bao, "Real time robust L1 tracker using accelerated proximal gradient approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1830–1837.

[26] Haibin Ling, "Online robust image alignment via iterative convex optimization," in *Computer Vision and Pattern Recognition*, 2012, pp. 1808–1814.

[27] Tianzhu Zhang, B Ghanem, Si Liu, and N Ahuja, "Robust visual tracking via multi-task sparse learning," in *Computer Vision and Pattern Recognition*, 2012, pp. 2042–2049.

[28] Caseiro Rui, Pedro Martins, and Jorge Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *European Conference on Computer Vision*, 2012, pp. 702–715.