VISUAL TRACKING WITH SPARSE CORRELATION FILTERS

Yanmei Dong, Min Yang, Mingtao Pei

Beijing Laboratory of Intelligent Information Technology
School of Computer Science, Beijing Institute of Technology, Beijing 100081, P.R. China
Email: {dongyanmei, yangminbit, peimt}@bit.edu.cn

ABSTRACT

Correlation filters have recently made significant improvements in visual object tracking on both efficiency and accuracy. In this paper, we propose a sparse correlation filter, which combines the effectiveness of sparse representation and the computational efficiency of correlation filters. The sparse representation is achieved through solving an ℓ_0 regularized least squares problem. The obtained sparse correlation filters are able to represent the essential information of the tracked target while being insensitive to noise. During tracking, the appearance of the target is modeled by a sparse correlation filter, and the filter is re-trained after tracking on each frame to adapt to the appearance changes of the target. The experimental results on the CVPR2013 Online Object Tracking Benchmark (OOTB) show the effectiveness of our sparse correlation filter-based tracker.

Index Terms— visual tracking, correlation filters, sparse representation, ℓ_0 regularization

1. INTRODUCTION

In visual object tracking methods based on correlation filters, the target appearance is modeled by correlation filters, and tracking is performed via convolution which becomes a simple element-wise multiplication in the Fourier domain. Due to the high computational efficiency, Correlation Filter-based Trackers (CFTs) have attracted considerable attention recently. Various trackers based on discriminative correlation filters have been proposed [1], and lots of these methods outperform state-of-the-art non-correlation filter-based trackers.

Although much improvement have been made in visual tracking, the performance of CFTs is still affected by appearance changes caused by variation of illumination, scaling, background clutter, and pose variations. CFTs employ only one image or a few images for training. The obtained filters might contain non-critical features about the visual object, and the presence of noise may result at the drifting or failure of tracking. To solve these problems, we utilize sparse representations in correlation filters for designing a robust and fast visual tracker. Sparse coding is able to represent the essential information of data while being insensitive to noise, and it has been widely used in visual tracking and generated state-of-the-art results[2, 3, 4, 5, 6].

In this paper, we propose a sparse correlation filter for visual object tracking. Our method employs ℓ_0 regularization to learn a sparse representation of correlation filters, taking advantages of both the robustness of sparse representations and the promising performance and high computational efficiency of correlation filters. During tracking, the tracked target appearance is modeled by a sparse correlation filter trained on image patches cropped from an initial position of the target at the first frame of the video. Then the filter is re-trained after tracking on each frame, from which we can obtain the new training data. To our best knowledge, we are the first one to propose sparse correlation filters using sparse representations, and apply it to solve visual tracking problems.

2. RELATED WORK

Different from other discriminative methods, correlation filter-based tracking methods regress all the circular-shifted variants of the input features to a target Gaussian function, so there is no need to sample a large number of negative and positive samples with hard-threshold. Bolme et al.[7] propose to learn a Minimum Output Sum of Squared Error (MOSSE) filter to model the appearance of the tracked object on gray-scale images. They can produce stable correlation filters even initialized with a single frame. With the use of correlation filters, the MOSSE tracker achieves high tracking efficiency with speed reaching several hundreds frames per second.

On the basic framework of MOSSE filter-based tracker, numerous improvements have been made later. For instance, by exploiting the circulant structure of the data matrix, Henriques et al. [8] apply correlation filters to kernel space, and propose the CSK tracker. Afterward, they extend the CSK method with HOG features, and propose a Kernelized Correlation Filter (KCF) and a Dual Correlation filter (DCF) with linear kernel[9]. To better represent the input data, color naming features, HOG features, features extracted from deep convolutional neural networks are employed to the correlation filters [10, 11, 12, 13]. And works [14, 15] are presented to handle scale variations on the object. To achieve successful tracking in handling long-term occlusion and out-of-view problem-

s, Ma et al.[16, 17, 18] adopt occlusion detection schemes for long-term tracking. These methods achieve superior performance compared with other state-of-the-art trackers.

In this work, we propose a spare correlation filter, and the most related works are [7, 9, 11] where correlation filters are presented firstly or improved using multi-channel features (eg. HOG). Compared to these correlation filters, our method considers the noise appearing in visual tracking, and we aim to design a robust correlation filter for visual tracking. In our method, sparse representation is utilized to learn a sparse correlation filter to represent the essential information of data, and the learned sparse correlation filter is insensitive to noise.

3. THE PROPOSED TRACKER

In correlation filter-based tracking methods, the target appearance is modeled by correlation filters trained on image patches extracted from the initial frame of a video, and tracking is performed via correlation over the filter and a search window in the next frame. The location corresponding to the maximum response of correlation results indicates the new target position. To successfully track the target in subsequent frames, the correlation filter is updated online according to that new position.

3.1. Sparse Correlation Filter

In order to construct a sparse representation of the tracked target with CFTs, we minimize the output sum of square error between the desire output and the observed output, and add ℓ_0 regularization to the filter. This minimization problem can be expressed by

$$\min_h \sum_i \parallel f_i \otimes h - g_i \parallel^2 + \lambda \parallel h \parallel_0, \tag{1}$$

where f_i is the i-th training image patch, h is the required correlation filter, g_i is the desired output of correlation which has a compact 2D Gaussian shaped peak centered on the target in f_i, $\otimes$ indicates correlation operation, and λ controls the sparsity of the filter.

It is notable that the original ℓ_0 regularized optimization is a NP hard problem. Consequently, we utilize a alternating optimization strategy with half-quadratic splitting [19] to obtain an approximation solution. In the first step, we rewrite the objective function (1) as

$$\min_h \sum_i \parallel f_i \otimes h - g_i \parallel^2 + \lambda c(h), \tag{2}$$

where $c(h) = \#\{p| \mid p \mid \neq 0\}$, p indicates the index of h, and $\#\{\}$ denotes counting operator. Next, a auxiliary variable v corresponding to h is introduced to simplify the optimization problem (2) as

$$\min_{h,v} \sum_i \parallel f_i \otimes h - g_i \parallel^2 + \lambda c(v) + \beta \parallel h - v \parallel^2, \tag{3}$$

where β is utilized to control the similarity between the auxiliary variable v and the filter h. Formula (3) approaches (2) when β is large enough.

Eq.(3) can be solved by alternatively minimizing h and v. In v minimization problem, the value of h is fixed with the result obtained from the previous iteration, and the same with v in the minimization problem of h.

3.1.1. Subproblem of minimizing h

By omitting the terms not involving h in Eq. (3), the h estimation subproblem is given by

$$\min_h \sum_i \parallel f_i \otimes h - g_i \parallel^2 + \beta \parallel h - v \parallel^2 . \tag{4}$$

After using Fast Fourier Transform (FFT) for speedup, the objective function above takes the form as

$$\min_H \sum_i \parallel F_i \odot H^* - G_i \parallel^2 + \beta \parallel H - V \parallel^2, \tag{5}$$

where $F_i = \mathcal{F}(f_i)$, $H = \mathcal{F}(h)$, $G_i = \mathcal{F}(g_i)$, $V = \mathcal{F}(v)$, in which $\mathcal{F}$ is the FFT operator, and symbol $\odot$ denotes element-wise multiplication while $*$ indicates the complex conjugate.

In the Fourier domain, correlation becomes element-wise multiplication, so we can estimate each element of the filter H separately,

$$\min_{H_p} \sum_i \mid F_{ip} H_p^* - G_{ip} \mid^2 + \beta \mid H_p - V_p \mid^2, \tag{6}$$

where p indexes the elements of H. This objective function is a real-valued, positive, and convex function of complex variables. Therefore, we can solve it via the method using in [7]. Setting the partial w.r.t. H_p equal to zero and solving the derivative, we can obtain the solution of H_p^*

$$H_p^* = \frac{\sum_i G_{ip} F_{ip}^* + \beta V_p^*}{\sum_i F_{ip} F_{ip}^* + \beta}. \tag{7}$$

Finally, the sparse correlation filter takes the form as

$$H = \frac{\sum_i F_i \odot G_i^* + \beta V}{\sum_i F_i \odot F_i^* + \beta}, \tag{8}$$

where multiplication and division are element-wise operations.

3.1.2. Subproblem of minimizing v

Omitting the terms not involving v in Eq.(3), the formula for minimizing v is defined as

$$\min_v \lambda c(v) + \beta \parallel h - v \parallel^2, \tag{9}$$

where $c(v)$ return the number of non-zero elements in v. This subproblem can actually be solved quickly because Eq. (9)

can be spatially decomposed where each element in v can be estimated individually,

$$\min_{v_p} \frac{\lambda}{\beta} H(|v_p|) + (h_p - v_p)^2, \tag{10}$$

where $H(|v_p|)$ is a binary function returning 1 if $|v_p| \neq 0$, and returning 0 otherwise. Solving Eq.(10), we can obtain the optima of v_p,

$$v_p = \begin{cases} 0, & \frac{\lambda}{\beta} \geq h_p^2 \\ h_p, & otherwise. \end{cases} \tag{11}$$

3.2. Filter Initialization and Updates

For a given video, an image patch, larger than the target region, is cropped from the initial position of the target at the first frame. And an initialized sparse correlation filter is obtained by training with this image patch.

During tracking, the object appearance often changes significantly due to partial or fully occlusion, deformation, fast motion, and the variety of rotation, scale, pose, illumination. Accordingly, it is a crucial part of visual tracking to update the filter online. Our sparse correlation filters are re-trained after tracking on each frame, and the terms about image patches in Eq. (8) for solving H_i are updated with a learning rate η after tracking on frame i:

$$H_i = \frac{A_i + \beta V}{B_i + \beta}, \tag{12}$$

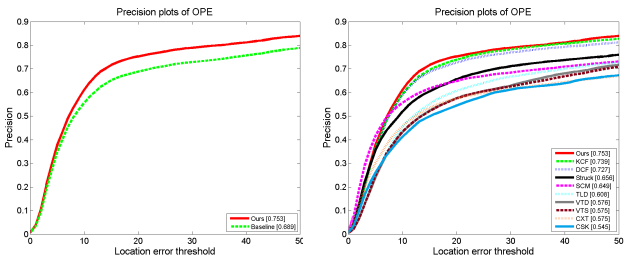
$$A_i = (1 - \eta) A_{i-1} + \eta F_i \odot G_i^*, \tag{13}$$

$$B_i = (1 - \eta) B_{i-1} + \eta F_i \odot F_i^*. \tag{14}$$

4. EXPERIMENTS

We implement our sparse correlation filter-based tracker in MATLAB on a desktop computer with an I7-2600 Intel 3.40 GHz CPU with 16 GB RAM. In the experiments, we presence a multi-channel sparse correlation filter with HOG features [20], using 9 orientation bins. Our sparse correlation filter requires a few more parameters, including ℓ_0 regularization parameter λ which is set to 0.2, parameter β controlling the similarity between the filter h and the auxiliary variable v, and parameter $padding$ fixed with 1.5 (denotes the tracked region is 1.5 times larger than the target region). β is automatically adapted in iterations starting from $\beta_0 = 0.02$ to $\beta_{max} = 10^5$ by a update rate $\kappa = 1.8$.

The pipeline for the tracker is intentionally simple, and does not including any additional information for failure detection. We first train a sparse correlation filter on some image patches obtained from the initial position of the target on the first frame. For a new frame, we detect over the image patch at the previous position, and update the target position with the position corresponding to the maximum value of the correlation results. Finally, we re-train a new model with image patches cropped from the new target position for tracking at a new frame.



(a) (b)

Fig. 1. Distance precision plots for all the 50 benchmark sequences using one-pass evaluation (OPE). (a) Distance precision plots of our tracker and a baseline tracker. (b) Distance precision plots of our tracker and state-of-the-art trackers, only the top ten ranked trackers and corresponding results are shown.

4.1. Sparsity Evaluation

In order to clearly demonstrate the effectiveness of the sparsity of correlation filters, we report the tracking results of our sparse correlation filter and a baseline tracker with the same parameter values. The distance precision plots for comparison are shown in Figure 1(a). From this figure, we can learn that our sparse correlation filter obviously outperforms the baseline tracker. Besides, our tracker performs favorably against the baseline under all the eleven video attributes annotated in the benchmark [21].

4.2. Overall Performance

Our trackers are evaluated on a famous visual object tracking dataset OOTB [21] that contains 50 video sequences with 51 targets. This dataset covers various challenging situations for visual tracking, including deformation, in-plane and out-of-plane rotation, partial occlusion, illumination variation, fast motion, etc.

The main performance criteria we used is precision curves. A frame may be successfully tracked if the location error between the predicted target center and the ground truth is smaller than a given threshold, ordinarily, this threshold is 20 pixels. Precision curves simply show the percentage of correctly tracked frames. Another popular choice is success curves, using bounding box overlap to evaluate the trackers. However, success curves heavily penalize trackers that do not track across scale, even if the target is tracked perfectly.

For comparison, we evaluate our trackers against the D-CF tracker, the KCF tracker, and all trackers summarized in

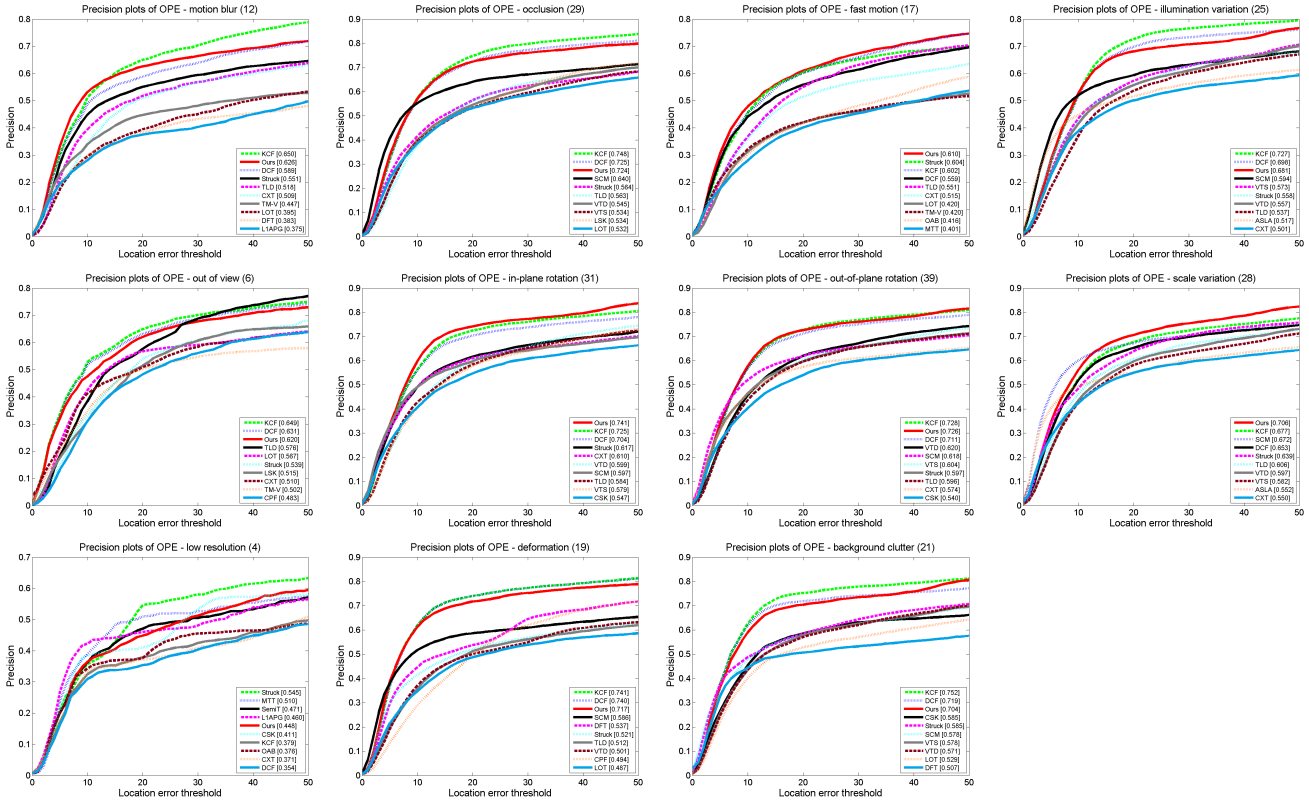


Fig. 2. Distance precision plots for eleven different tracking challenging attributes: motion blur, occlusion, fast motion, illumination variation, out of view, in-plane rotation, out-of-plane rotation, scale variation, low resolution, deformation, and background clutter. Our proposed tracker performs favorably against most trackers with these attributes.

[21]. Figure 1(b) shows the results under one-pass evaluation (OPE) using distance precision rate on all the 50 benchmark sequences. Overall, the proposed algorithm performs better than other methods, and achieves an distance precision rate of 75.3% while operating at a speed about 10 frames per second.

Further evaluations are performed to analyze the performance of sparse correlation filters under different video attributes. Figure 2 shows the OPE for all the eleven video attributes, including motion blur, occlusion, fast motion, illumination variation, out of view, in-plane rotation, out-of-plane rotation, scale variation, low resolution, deformation, and background clutter. From Figure 3, we can observe that our method performs favorably against other trackers.

5. CONCLUSION

In this paper, we have proposed a sparse correlation filter for visual object tracking by exploiting the sparse representation of the target. A minimization problem for sparse correlation filters is obtained through applying ℓ_0 regularization to the conventional correlation filter formula. To solve this optimal problem, we have introduced a auxiliary variable and utilized a alternating optimization strategy with half-quadratic splitting. During object tracking, we employ our new sparse correlation filter to model the appearance of the target, and the filter is re-trained after tracking on every frame. The experimental results on the OOTB dataset have shown that the proposed method outperforms other state-of-the-art methods, and the comparison with a baseline tracker has demonstrated the effectiveness of our sparsity on the filter.

6. REFERENCES

[1] Zhe Chen, Zhibin Hong, and Dacheng Tao, "An experimental survey on correlation filter-based tracking," *arXiv preprint arXiv:1509.05520*, 2015.

[2] Yuwei Wu, Bo Ma, Min Yang, Jian Zhang, and Yunde Jia, "Metric learning based structural appearance model for robust visual tracking," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 5, pp. 865–877, 2014.

[3] Min Yang, Yuwei Wu, Mingtao Pei, Bo Ma, and Yunde Jia, "Online discriminative tracking with active example selection," *Circuits and Systems for Video Technology, IEEE Transactions on, in press*, 2015.

[4] Yuwei Wu, Mingtao Pei, Min Yang, Junsong Yuan, and Yunde Jia, "Robust discriminative tracking via landmark-based label propagation," *Image Processing, IEEE Transactions on*, vol. 24, no. 5, pp. 1510–1523, 2015.

[5] Bo Ma, Jianbing Shen, Yangbiao Liu, Hongwei Hu, Ling Shao, and Xuelong Li, "Visual tracking using strong classifier and structural local sparse descriptors," *Multimedia, IEEE Transactions on*, vol. 17, no. 10, pp. 1818–1828, 2015.

[6] Xiangyuan Lan, Andy J Ma, Pong C Yuen, and Rama Chellappa, "Joint sparse representation and robust feature-level fusion for multi-cue visual tracking," *Image Processing, IEEE Transactions on*, vol. 24, no. 12, pp. 5826–5841, 2015.

[7] David S Bolme, J Ross Beveridge, Bruce Draper, Yui Man Lui, et al., "Visual object tracking using adaptive correlation filters," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2544–2550.

[8] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision–ECCV 2012*, pp. 702–715. Springer, 2012.

[9] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista, "High-speed tracking with kernelized correlation filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 3, pp. 583–596, 2015.

[10] Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, and Joost van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1090–1097.

[11] Hamed Kiani Galoogahi, Terence Sim, and Simon Lucey, "Multi-channel correlation filters," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3072–3079.

[12] Yang Li and Jianke Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Computer Vision-ECCV 2014 Workshops*. Springer, 2014, pp. 254–265.

[13] Chao Ma, Jia-Bin Huang, Xiaokang Yang, and Ming-Hsuan Yang, "Hierarchical convolutional features for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3074–3082.

[14] Qianyun Du, Zhao-quan Cai, Hao Liu, and Zhu Liang Yu, "A rotation adaptive correlation filter for robust tracking," in *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1035–1038.

[15] Lei Zhang, Yanjie Wang, Honghai Sun, Zhijun Yao, and Shuwen He, "Robust visual correlation tracking," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[16] Chao Ma, Xiaokang Yang, Chongyang Zhang, and Ming-Hsuan Yang, "Long-term correlation tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5388–5396.

[17] Liyang Yu, Chunxiao Fan, and Yue Ming, "A visual tracker based on improved kernel correlation filter," in *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*. ACM, 2015, p. 60.

[18] Zhibin Hong, Zhe Chen, Chaohui Wang, Xue Mei, Danil Prokhorov, and Dacheng Tao, "Multi-store tracker (muster): a cognitive psychology inspired approach to object tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 749–758.

[19] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia, "Image smoothing via l0 gradient minimization," in *ACM Transactions on Graphics (TOG)*. ACM, 2011, vol. 30, p. 174.

[20] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, 2010.

[21] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Online object tracking: A benchmark," in *Computer vision and pattern recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2411–2418.