Signal Processing: Image Communication 62 (2018) 117–128

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

# Context multi-task visual object tracking via guided filter

Yong Wang [a], Xinbin Luo [b,*], Shan Fu [b], Shiqiang Hu [c]

[a] School of Electrical Engineering and Computer Science, University of Ottawa, Canada
[b] School of Electronic information and Electrical Engineering, Shanghai Jiao Tong University, China
[c] School of Aeronautics and Astronautics, Shanghai Jiao Tong University, China

ARTICLE INFO

*Keywords:*
Visual object tracking
Context information
Multi-task sparse learning
Guided filter
Alternating direction method of multipliers

ABSTRACT

In this paper, we formulate particle filter based tracking as a multi-task sparse learning problem that exploits context information. The target and context information which modeled as linear combinations of principal component analysis (PCA) basis is formed as dictionary templates. The target is treated as the guidance and the sampling examples are filtered depending on the similarity between the target and each input. The edge preserving smoothing property of the guided filter is a key factor for object tracking. First, valuable candidates can be selected and the inaccurate candidates become blurry. Therefore, the guided filter can help to distinguish the target from numerous candidates easily. Second, partial occluded target can be recovered by the filtering process via the guidance image. Thus, the drifting problem can be alleviated. Then multi-task sparse learning is employed to learn the target and context information. The proposed learning problem is efficiently solved using an alternating direction method of multipliers (ADMM) method that yields a sequence of closed form updates. We test our tracker on challenging video sequences that involve drastic illumination changes, large pose variations, and heavy occlusions. Experimental results show that our tracker consistently outperforms state-of-the-art trackers.



## 1. Introduction

Visual object tracking plays a crucial role in computer vision including intelligent human computer interaction, video surveillance, action recognition to name a few. Although a lot of tracking algorithms have been proposed, it still remains a challenging problem for complex scenarios, e.g., illumination variation, background clutter, fast motion and so on (see Fig. 1).

A general method used to construct a robust tracking system involves two key components, an appearance model to evaluate the likelihood of each candidate state and select the best candidate as the target; a motion model, e.g., particle or Kalman filter, which aims to forecast the likely movements of the target over time to supply the tracker with a number of candidate states.

In this study, we develop an effective appearance model for robust object tracking, which deals with the first of the two components. We consider two critical factors. The first factor is how objects are represented. Several features have been used in object tracking and detection, e.g., intensity [1–4], color [5,6], texture [7], Haar feature [8–10], superpixel feature [11,12], and subspace representation [13]. Subspace representation maintains holistic appearance information and provides a more compact notion of the target than the other features. We model the tracking target with principal component analysis (PCA) basis vectors similar to [13]. Linear combination of the basis vectors can well represent uncorrupted samples. Also, context information which is used in visual tracking [14–19] is employed to improve the robustness of our method. Furthermore, the guided filter which has showed good tracking results [20] is adopted in our work. We treat target as the guidance image. Thus, the intrinsic characteristics of the target are preserved and the occlusion and illumination effect are alleviated. With the aid of guided filter, the tracker can distinguish the target from the background easily.

Second, previous representation schemes mainly focus on the $l_1$ technique. However, the relationship between the candidates is not considered. Multi-task sparse learning [21,22] is proposed, which aims to consider the representation of particles jointly sparsity. The decomposition form of multi-task sparse learning can be used to simultaneously capture a common set of features among relevant tasks and identifies outlier tasks [23–25]. In this paper, the representation model is decomposed into two collaborative parts, a row-sparse matrix which corresponding to the overlapping features and an elementwise sparse

* Corresponding author.
*E-mail address:* losinbin@sjtu.edu.cn (X. Luo).

https://doi.org/10.1016/j.image.2017.12.008
Received 9 August 2017; Received in revised form 18 November 2017; Accepted 19 December 2017
Available online 1 January 2018
0923-5965/

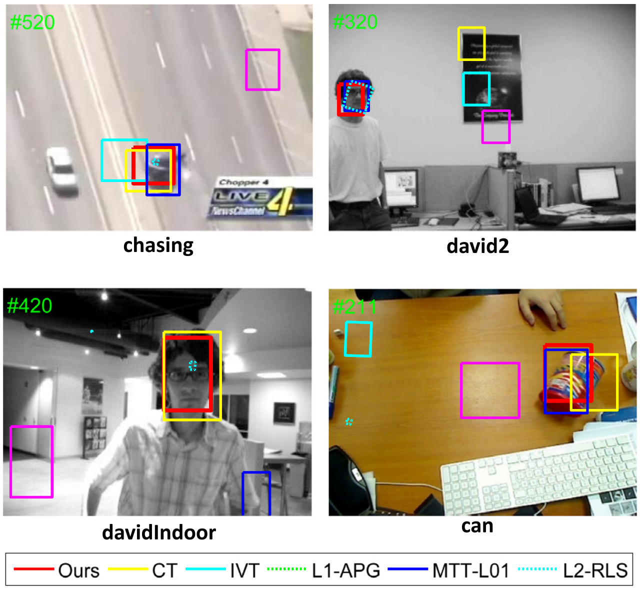


**Fig. 1.** Challenge scenarios encountered in visual object tracking. Existing trackers are not able to effectively deal with rotation, illumination variation, background clutter and fast motion, while our method gives more robust results.

matrix which corresponding to the non-shared features or outliers. Alternating direction method of multipliers (ADMM) [26] which guarantees convergence of this model is employed to solve the model.

Inspired by the above work, we develop a computationally efficient, multi-task sparse learning tracker that exploits context information. The contributions of this paper are threefold. (1) We integrate context information into subspace learning method, which makes use of context information for robust tracking performance. (2) The guided filter is employed to alleviate the effect of illumination and background clutter. The guided filter preserves the edge information. That is, most of the characteristics of the target can be retained. (3) The decomposition model is employed to separate the common part and outliers in the candidates. This model enables our tracker to choose the appropriate candidate effectively. ADMM is adopted to solve the model effectively and robustly.

A preliminary study is presented in [27]. In this paper, we provide more comprehensive analysis and descriptions of our method with detailed implementation and thorough experimental comparison. We compare with 9 state-of-the-art methods on 45 sequences. In addition, full analysis is presented to demonstrate the effectiveness and robustness of the proposed tracking algorithm.

The remainder of the paper is organized as follows: In the next section, related work is summarized. After that, the tracking algorithm is given in Section 3. Experimental results are reported in Section 4. Finally, conclusions are drawn in Section 5.

## 2. Related work

Visual object tracking has been studied extensively by many researchers over the years due to its importance. While a comprehensive review of the tracking methods is beyond the scope of the paper, please refer to [28,29] for a survey, and also to [30,31] for some empirical comparisons.

In general, tracking algorithms can be categorized into generative and discriminative approaches. Generative tracking methods learn an appearance model to represent the target object and then use it to search for the target location with minimal reconstruction error [1,2,13,20,32,33]. Sparse representation has been used in the object tracking where an object is modeled by a sparse linear combination of target and trivial templates [1]. Accelerated proximal gradient (APG) approach is used to improve the tracking performance since the computation complexity of $l_1$ method [2]. Subspace learning based method is first introduced in [13] to adaptively update the templates and achieves excellent results. Guided filter is first introduced into visual tracking to improve the image quality [20].

Discriminative algorithms formulate visual tracking as a binary classification problem, which seek the optimal decision boundary for separating the target from the background. Various algorithms have been proposed with demonstrated success, e.g., [7–10,34], to name a few. An online boosting algorithm is proposed to select features for tracking [8]. Boosting method has been used for object tracking [7] by combining weak classifiers with pixel based features within the target and background regions with the on-center off-surround principle. However, these discriminative algorithms [7,8] utilize only one positive sample (i.e., the tracking result in the current frame) and multiple negative samples when updating the classifier. If the object location detected by the current classifier is not precise, the positive sample will be noisy and result in a suboptimal classifier update. Consequently, errors will be accumulated and cause tracking drift or failure. Multiple instance learning (MIL) [9] is employed and equipped with a Haar feature pool to overcome the problem mentioned above. An appearance model based on non-adaptive random projections that preserves the structure of original image space is proposed for robust tracking [10].

Multi-task sparse learning [22] has been successfully applied to image classification [21] and image annotation [35]. A multi-task joint covariate selection model is used to classify a query image using multiple features from a set of training images, and a class-level joint sparsity regularization is imposed on representation coefficients [21]. Multi-task sparse learning approach is proposed to jointly learn the particle representations for robust object tracking [3]. Instead of handling particles separately, particle representations are jointly learned to exploit similarities among particles. Similarly, low rank constraint is imposed on the joint optimization of the candidate groups [4]. Compared to the $l_1$ method [1] that pursues the sparse representation independently, multi-task sparse learning achieves more robust performance by exploiting the interdependency between tasks. Structural Sparse tracker [12] exploits the relationship among candidates and their local parts to learn their sparse representations jointly. A circulant sparse appearance model for visual tracking is proposed in [36]. The HOG feature is employed and the optimization can be solved efficiently due to the circulant structure.

Besides, context information has been successfully applied in object recognition [37], saliency detection [36], object classification [38], object detection [39] and visual object tracking [14–19]. The improved performance of these tracking methods is attributed to the use of context information in determining the object location. Context information is considered in the scene for robust object tracking [14]. A number of auxiliary object are integrated into the tracking process. A new tracker has been proposed that combines the concept of context and structure for object tracking [15]. In [16], the authors exploit context information and object appearance to improve tracking results. Spatio-temporal relationships between the object and its local context are considered in [17]. The statistical correlation between the features from the target and its surrounding regions is formulated to a confidence map which is maximized to an object location. In [18], the tracker locates the object under heavy occlusion by context information. In our point of view, if the target representation incorporates appropriate background information, it in turn prevents the tracker to drift from the target into the background.

Our work is closely related to [20], while we make two improvements: context information and multi-task sparse learning. All the previous work suggests that target appearance model can be improved context information and multi-task sparse learning. Thus, we are interested in exploring the related information shared between candidates in order to obtain improved performance. We present our proposed algorithm in the following section.