Signal Processing: *Image Communication* 37 (2015) 1–18

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Object tracking using discriminative sparse appearance model

Dandan Huang, Yi Sun *

Department of Electronic Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

ARTICLE INFO

Article history:
Received 27 January 2015
Received in revised form
13 May 2015
Accepted 29 June 2015
Available online 16 July 2015

Keywords:
Visual tracking
Sparse representation
Dictionary learning
Adaptive update
Bayesian inference framework

ABSTRACT

Object tracking based on sparse representation formulates tracking as searching the candidate with minimal reconstruction error in target template subspace. The key problem lies in modeling the target robustly to vary appearances. The appearance model in most sparsity-based trackers has two main problems. The first is that global structural information and local features are insufficiently combined because the appearance is modeled separately by holistic and local sparse representations. The second problem is that the discriminative information between the target and the background is not fully utilized because the background is rarely considered in modeling. In this study, we develop a robust visual tracking algorithm by modeling the target as a model for discriminative sparse appearance. A discriminative dictionary is trained from the local target patches and the background. The patches display the local features while their position distribution implies the global structure of the target. Thus, the learned dictionary can fully represent the target. The incorporation of the background into dictionary learning also enhances its discriminative capability. Upon modeling the target as a sparse coding histogram based on this learned dictionary, our tracker is embedded into a Bayesian state inference framework to locate a target. We also present a model update scheme in which the update rate is adjusted automatically. In conjunction with the update strategy, the proposed tracker can handle occlusion and alleviate drifting. Comparative results on challenging benchmark image sequences show that the tracking method performs favorably against several state-of-the-art algorithms.

1. Introduction

Object tracking is important in computer vision applications such as video surveillance, traffic monitoring, and sports analysis. Thus, various tracking algorithms have been proposed in the literature. Comprehensive reviews are presented in [1] and [2]. Nonetheless, robust tracking remains a challenging problem because of the variation in target appearance as a result of partial occlusions, illumination and pose changes, clutter scenes, and scale variation. Samples are shown in Fig. 1.

Two types of approaches can generally be employed to achieve robust tracking. One is the generative method, which builds an effective target appearance model to adapt to the variation. Target tracking is formulated as determining the region that is highly similar to the target model in image sequences. The target appearance model is usually constructed according to target templates [3–7] or their subspace models [8,9]. In addition, superpixel-based [10], coupled-layer [11], color-based [12], and parts-based models [13,14] are also used to model the target appearance. This method captures changes in appearance by updating the model automatically [4,9,15]. The other tracking approach is the discriminative method, which treats object tracking as a binary classification problem. That is, it directly distinguishes the target from the background. This method uses the features extracted from both the target and the background,

* Corresponding author.
E-mail address: lslwf@dlut.edu.cn (Y. Sun).

http://dx.doi.org/10.1016/j.image.2015.06.012
0923-5965/

Fig. 1. Frames from *david*, *car11*, *football*, *faceocc2*, *singer1* and *jumping* sequence, respectively. The ground truth target regions are bounded by red rectangles. The main challenges are shown at the bottom of the images in red fonts.

and it quite depends on classifier training. Unsupervised learning [16–20] is applied to train the classifier, but this method results in potential drifting because no supervisory mechanism has been built into classifier updates. To address this problem, semi-supervised [21–23] and multiple-instance learning [24] are proposed. In addition, P-N learning [25] learns the classifier that focuses on exploiting the structure of the training samples. An incremental learner with a linear support vector machine (SVM) and least square loss [26] represents the target as a binary code and separates the target from its surroundings.

Recent studies have confirmed that sparse representation is effective in computer vision application [27–29]. Therefore, several tracking algorithms based on sparse representation have been presented. The sparsity-based tracking method firstly trains a dictionary that is also known as a set of basis vectors. Then, the target appearance is modeled on the basis of the learned dictionary. During tracking, the candidates are sparsely represented in the space spanned by the dictionary. Finally, the candidate with minimal reconstruction error is regarded as the tracking result. Sparsity-based tracking has progressed significantly, but an effective method for building a target appearance model remains a question worth of studying. The existing sparsity-based appearance model has two main problems. One is that the appearance model is individually represented using holistic or local templates; as a result, the global structural information cannot be sufficiently combined with local features. This problem may induce tracking failure in the event of occlusion or rapid changes in appearance. The other problem is that the background is rarely considered in target appearance modeling; hence, the discriminative information between the target and the background is not fully utilized.

In the current paper, we propose an object tracking method based on a discriminative sparse appearance model and a dynamic template update. First, we learn a dictionary from both the target region and the background around the target. Each basis of the dictionary is a vectorized local image patch that is obtained from the target region. It is worth pointing out that the bases in our dictionary are not disorderly but ordered specifically. Then, the target appearance is modeled as a sparse coding histogram based on the learned dictionary. After that the tracking is embedded into a Bayesian inference framework. And the observation likelihood is modified into a similarity coefficient, which is defined to measure the similarity between the target model and the candidate. A strategy for model updating is also proposed to limit the adverse effects caused by appearance changes.

The details or main advantages of our method are summarized as follows.

(1) There are two advantages about our dictionary. On the one hand, the incorporation of the background into learning enables the dictionary to enhance its discriminative capability. On the other hand, the bases in the constructed dictionary are local image patches selected from the target region; thus, the subspace spanned by the dictionary is compact. This compact subspace ensures that the targets can be represented well. Overall, our dictionary considers both expression capability and discriminative strength; therefore, it is more suitable for object tracking task.

(2) Target appearance is modeled as a sparse coding histogram of specified image patches. The positions of these specified patches are the same as the bases that construct the dictionary. These patches are chosen because they not only represent the local features of the target but also distinguish the target from the background. The spatial position distribution of these patches also implies the global structure of the target. As the sparse coding histogram of these significant patches, our appearance model combines the local features and structure of the