Image and Vision Computing 32 (2014) 930–939

Contents lists available at ScienceDirect

Image and Vision Computing

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

Robust object tracking using least absolute deviation

Jingyu Yan, Fuxiang Wang *, Xianbin Cao, Jun Zhang

School of Electronic and Information Engineering, Beihang University, Beijing, China

ARTICLE INFO

Article history:
Received 6 August 2013
Received in revised form 10 March 2014
Accepted 28 August 2014
Available online 6 September 2014

Keywords:
Object tracking
Sparse representation
Least absolute deviation

ABSTRACT

Recently, sparse representation has been applied to object tracking, where each candidate target is approximately represented as a sparse linear combination of target templates. In this paper, we present a new tracking algorithm, which is faster and more robust than other tracking algorithms, based on sparse representation. First, with an analysis of many typical tracking examples with various degrees of corruption, we model the corruption as a Laplacian distribution. Then, a LAD–Lasso optimisation model is proposed based on Bayesian Maximum A Posteriori (MAP) estimation theory. Compared with L1 Tracker and APG-L1 Tracker, the number of optimisation variables is reduced greatly; it is equal to the number of target templates, regardless of the dimensions of the feature. Finally, we use the Alternating Direction Method of Multipliers (ADMM) to solve the proposed optimisation problem. Experiments on some challenging sequences demonstrate that our proposed method performs better than the state-of-the-art methods in terms of accuracy and robustness.

1. Introduction

Object tracking is an important component of many surveillance systems, such as transport systems (e.g., road traffic, airports and harbours), public spaces (e.g., shopping malls and parks), industrial environments, low-altitude rescue and military establishments. More efficient and robust object tracking remains a challenge due to the issues of image noise, complex object motion, partial or full occlusions, drastic illumination and pose changes [1].

Methods based on particle filter are widely used for tracking. Under the particle filter framework, different image features such as colour [3,4,7], shape [2,13] and image structure [7,15] can be used to represent the appearance of object.

Recently, there has been an increased interest in sparse representation and its applications in the field of computer vision. Under the particle filter framework, Xue Mei et al. [8,9] proposed a tracking method based on sparse representation, which was named the L1 Tracker. To find the tracking target in a frame, each target candidate (particle sample) is approximately expressed as a sparse linear combination of some target templates and trivial templates. The sparse representation coefficients of a target candidate are calculated by solving a L1-regularised least squares problem, which is high computational cost due to the high dimension of trivial templates.

To further accelerate the L1 Tracker, Hanxi Li et al. [16] use the orthogonal matching pursuit (OMP) algorithm to search for a sparse solution. To reduce the number of particle samples that need to participate in solving the optimisation problems, Xue Mei et al. [10] improve L1 Tracker with a minimal error bounding strategy called the BPR-L1 Tracker. The APG-L1 Tracker proposed by Chenglong Bao et al. [24] used an accelerated proximal gradient (APG) approach to solve a new L1-norm related problem that added one term to control the energy of trivial templates. By regularizing the representation problem to enforce joint sparsity and learning the particle representations together, Tianzhu Zhang et al. [34] propose a computationally efficient multi-task sparse learning method to mine correlations among different tasks to obtain better tracking results than learning each task individually. The linear representation in their later work [37] incorporates background templates in the dictionary to discriminate the target from the background better and casts the tracking problem as an efficient low-rank matrix learning problem.

To further accelerate the L1 Tracker, Hanxi Li et al. [16] use the orthogonal matching pursuit (OMP) algorithm to search for a sparse solution. To reduce the number of particle samples that need to participate in solving the optimisation problems, Xue Mei et al. [10] improve the computation speed of L1 Tracker with a minimal error bounding strategy called the BPR-L1 Tracker. The APG-L1 Tracker proposed by Chenglong Bao et al. [24] used an accelerated proximal gradient (APG) approach to solve a new L1-norm related problem that added one term to control the energy of trivial templates. By enforcing joint sparsity, Tianzhu Zhang et al. [34] propose a computationally efficient multi-task sparse learning method, which learns the particle representation solutions of all particle samples together.

* Corresponding author at: School of Electronic and Information Engineering, Beijing University of Aeronautics and Astronautics, Xueyuan Road No. 37, Haidian District, Beijing, China. Tel.: +86 10 82316281.
E-mail addresses: yan_jy@163.com (J. Yan), wangfx@buaa.edu.cn (F. Wang), xbcao@buaa.edu.cn (X. Cao), buaazhangjun@vip.sina.com (J. Zhang).

http://dx.doi.org/10.1016/j.imavis.2014.08.008
0262-8856/

Some work based on sparse representation is devoted to improve the robustness of tracking. To alleviate the accumulation of errors during the self-updating, Baiyang Liu et al. [31] use a static sparse dictionary and a dynamically online updated basis distribution to model the target appearance. In order to deal with the challenge of drastic appearance change, Zhong Wei et al. [32] propose an appearance model exploiting both holistic templates and local representation. Jia Xu et al. [33] develop an appearance model, which exploits both partial information and spatial information of the target based on a novel alignment-pooling method. To discriminate the target from the background, Tianzhu Zhang et al. [37] incorporates background templates into the dictionary of sparse representation and reformulates the tracking problem as an efficient low-rank matrix learning problem.

In the view of the model of the representation error, LSS [36] assumes that the representation error follows the Gaussian–Laplacian distribution. Dong Wang et al. [35,36] use classic principal component analysis (PCA) to learn effective appearance model. It needs to be stressed that there is no sparsity constraint on the representation coefficients in [35,36]. Consequently LSS [36] is not in the framework of sparse representation.

In this paper, we propose a new tracking method under the framework of the L1 Tracker that can work more quickly and robustly. Our main contributions include:

1) The representation error is modelled as a random variable following a Laplacian distribution. The representation error, which indicates corruption or noise, is random and unknown in advance. Thus, accurately modelling the representation error is a key to the robustness of tracking. After an elaborate analysis of the distribution of the corruption, we find that the distribution of the corruption is characterised by one spike and a long-tail, so we model the representation error as a Laplacian distribution.
2) Based on the Laplacian representation error model and a sparseness-promoting prior of the representation vector, we derive our new LAD–Lasso model with a Bayesian Maximum A Posteriori (MAP) estimate. The number of optimisation variables in our new model is equal to the number of target templates, regardless of the dimensions of the feature. Thus, the computation cost can be reduced greatly compared with L1 Tracker and APG-L1 Tracker.
3) After reformulating our proposed optimisation model, we use Alternating Direction Method of Multipliers (ADMM) to solve our proposed nonsmooth optimisation problem.

We name our new method the LAD Tracker (Least Absolute Deviation). Experiments on challenging video sequences demonstrate our method performs well in computation speed and robustness.

This paper is organised as follows: In Section 2, we briefly review the basic idea of trackers based on sparse representation. Section 3 introduces our LAD Tracker in detail. In Section 4, we make a theoretical analysis of the robustness and computation cost, compared with other trackers based on sparse representation. In Section 5, we demonstrate the performances of the LAD Tracker through numerous experiments. The conclusion is made in Section 6.

2. Trackers based on sparse representation

In this section, we will briefly introduce the framework of trackers based on sparse representation. John Wright and Yi Ma et al. [20] addressed the problem of human face recognition via computing sparse linear representations with regard to a dictionary of different human faces. Then, Xue Mei et al. [8] extend the application based on sparse representation to tracking named the L1 Tracker. There are many later works that improve on this method [10,16,24]. All of these related trackers inherit the framework of the L1 Tracker.

Generally, there are three main components in a tracking system [6]: the motion model, image representation and appearance model. L1 Tracker uses Markov process along with a Gaussian sampling strategy in a particle filter as its motion model, downsampling the image as its image representation and an adaptive updating template set as its appearance model. Both the particle filter and downsampling image allow L1 Tracker to adapt itself to scale variation. For more details about particle filters, please refer to the tutorial on particle filters [12, 13]. The dynamic appearance model and model updating are related to the previous work on the appearance adaptive particle filter (AAPF) tracker [21].

The basic idea of L1 Tracker is that each particle sample in a given frame can be approximately expressed with a sparse linear combination of some target templates and trivial templates. The target template is dynamically updated when a new target candidate appears to be different from the template set. The process of L1 Tracker is illustrated in Fig. 1.

In the first frame, L1 Tracker chooses the first target template manually. Then, the other m templates are cropped by randomly moving only a few pixels around the first template, as shown in Fig. 1. These m templates form a positive template set. Each template is resized to

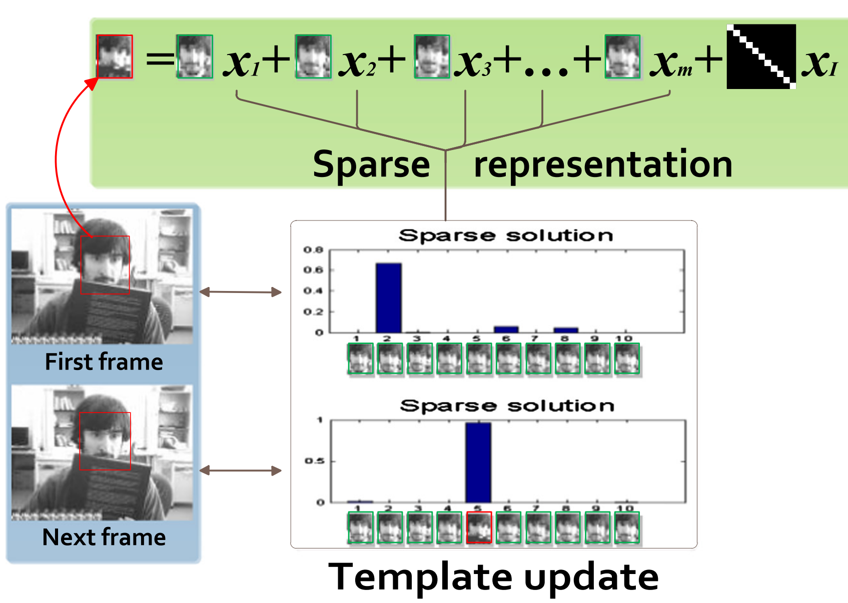

Fig. 1. The basic concept of the L1 Tracker.