Neurocomputing 154 (2015) 41–49

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-task l_0 gradient minimization for visual tracking

Hongwei Hu, Bo Ma*, Yunde Jia

Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form
27 October 2014
Accepted 2 December 2014
Communicated by Jinhui Tang
Available online 17 December 2014

Keywords:
l_0-norm minimization
l_1-norm minimization
Visual tracking
Multi-task
Sparse coding

ABSTRACT

In most object tracking algorithms based on sparse representation, the optimization problem is often formulated as an l_1 or l_2 minimization problem, because its primal l_0-norm minimization problem is NP-hard. In this paper, a visual tracking method is proposed based upon l_0-norm minimization which directly seeks solution to the primal l_0 problem. To avoid solving a large number of l_0 minimization problems, we introduce to encode all samples simultaneously in a multi-task manner, which means that the number of minimization problem to be solved is only one, and an algorithm is presented to solve the minimization problem. Our tracking algorithm is then implemented under the framework of particle filter. Experiments on different challenging video sequences demonstrate that our method can achieve robust tracking results.

1. Introduction

Visual tracking is still an active research topic in computer vision, although it has been studied for decades. Lots of object tracking algorithms have been proposed in recent decades, and surveys can be found in [1–3]. However, visual tracking remains a difficult problem to solve, since target during tracking suffers background clutter, abrupt motion and appearance variations such as illumination changes, partial or global occlusions and shape deformation [4].

Generally, visual tracking approaches can be classified into generative method and discriminative method. Generative tracking methods [5–11] aim to track a target by searching for the most similar region in the following frames with target templates. Kwon and Lee [5] decomposed observation model and motion model into several basic models to handle different appearances and motions of object. Adam et al. [6] represented an object template by multiple image patches and presented a robust tracker using integral histogram. Zhang et al. [8] extracted features on target appearance in compressive sensing domain for fast compressive tracking. Zhang et al. [9] proposed a multi-task approach for visual tracking under the framework of particle filter based on sparse learning. Oron et al. [10] handled rigid and nonrigid deformation of target appearance during tracking by a joint appearance and spatial configuration of pixels model. To represent object appearance effectively, Sevilla-Lara and Learned-Miller [11] employed the distribution of pixels with multiple layers. Although much success has been made by generative tracking methods, it still remains some problems. For example, a large number of training samples must be collected to build a robust appearance model, and drift may occur when target could not be reconstructed by templates well because of appearance changes. Discriminative methods [12–17] train a classifier, which distinguish foreground from background, with a set of training examples. Babenko et al. [12] alleviated the drift by integrated multiple instance learning method. Wang et al. [13] segmented image into superpixels and calculated confidence map by these superpixels on both foreground and background. Fan et al. [14] proposed tracking target by the spatial attention regions calculated from foreground and background regions. Kalal et al. [16,17] developed a P-N learning method to train a binary classifier and proposed a long-term tracking method by decomposing it into tracking, learning, and detection. Discriminative methods regard tracking as a binary classification problem. Several tracking algorithms [18,19] are presented exploiting the advantages of the both. Zhong et al. [18] proposed a sparsity-based collaborative model which applied both holistic templates and local representations. Yu et al. [19] used the hybrid discriminative generative model which described the object appearance and handled appearance variations using an online support vector machine.

Recently, sparse coding based visual tracking methods have attracted much attention in computer vision community. It was inspired by the success of sparse coding in face recognition [20] and showed promising tracking results [21]. Sparse coding based tracking methods could be classified into global sparse appearance based methods and local sparse appearance methods according to the appearance model established. Global sparse appearance based methods consider only global appearance of interesting

* Corresponding author.
E-mail addresses: huhongwei@bit.edu.cn (H. Hu), bma000@bit.edu.cn (B. Ma), jiayunde@bit.edu.cn (Y. Jia).

http://dx.doi.org/10.1016/j.neucom.2014.12.021

target, which may fail to track a target when its local appearance changes greatly, especially in the case of partial occlusion. As a pioneering work, Mei and Ling [22] introduced sparse representation into visual tracking. Under the framework of particle filter, they proposed a tracking algorithm by regarding visual tracking as a sparse reconstruction problem. By solving an l_1-regularized least squares problem, the sparse representation was achieved. However, they had to solve the same number of l_1-norm minimization problems as that of particles, which led to expensive computational cost. Hence, Bao et al. [23] introduced the Accelerated Proximal Gradient (APG) approach to improve the computing speed. Wang et al. [24] supposed that noise term was Gaussian–Laplacian distributed and proposed a least soft-threshold squares algorithm, which was actually a variant of l_1 sparse coding. Nonetheless, the real distribution of noise was unknown. Local sparse appearance based methods divide target templates into blocks or patches, which can alleviate the influence of partial occlusion. Liu et al. [25] argued that a local sparse appearance model with K-selection ought to be valid. The target was located by performing mean-shift [26] on voting confidence map constructed from reconstruction errors. However, this method did not allow for the spatial relationship of local image patches, which might cause drift. Motivated by Liu's work, Wang et al. [27] and Jia et al. [28] proposed a discriminative and adaptive tracking approach using local sparse appearance model. Different from other discriminative tracking methods, Wang et al. [27] learned sparse codes from raw image patches other than a pool of features. Jia et al. [28] exploited a novel alignment-pooling method using both partial information and spatial information of target. To model local appearance, Wang et al. [29] even learned visual prior from generic real-world images and transferred it into local sparse representation. Wu et al. [30] seek for an appropriate metric in the feature space of sparse codes, and tracking target with metric learning based structural appearance model. Hu et al. [31] propose to track object by nonlinear learning based on local coordinate coding.

In most object tracking algorithms based on sparse coding, the optimization problem is often formulated as an l_1 or l_2 minimization problem or the combination of these two norms [29], because its primal l_0-norm minimization problem is NP-hard. Moreover, l_1-norm minimization is the optimal convex approximation of l_0-norm minimization, and l_2-norm is effective to prevent over-fit. Therefore, people pay more attention on l_1 and l_2-norm problem and ignore the primal l_0-norm problem. Since proponents of sparsity based tracking are convinced that sparsity can address visual tracking effectively, this paper directly seeks solution to the primal l_0-norm minimization for sparse coding rather than its simplified l_1-norm version. In [32], Xu et al. presented a new image editing method which performed image smoothing and natural image deblurring via l_0 gradient minimization. In this paper, we use the l_0-norm minimization scheme to perform sparse coding, and a comparison experiment between l_0 and l_1 sparse coding shows that l_0 sparse coding is benefit for visual tracking than l_1 sparse coding. The visual tracking then can be formulated as an l_0-norm minimization problem. A very simple iterative method is presented by introducing an auxiliary variable to solve l_0-norm minimization problem approximately, which could yield the sparse code effectively.

Generally, sparse coding based tracking approaches are under the framework of particle filter. The number of optimization problems is proportional to the number of particles. To improve the computational efficiency, Zhang et al. [9] and Zhuang et al. [33] formulate object tracking as a multi-task sparse learning problem which means that the tracking problem can be solved with a single optimization problem, no matter how many the particles are. In inspired by their works, we further formulate these l_0-norm minimization problems using multi-task manner. As in [33], under the assumption that similar samples should have the similar sparse coefficients, a Laplacian term is also included. A simple and efficient algorithm is introduced to solve the proposed multi-task version of l_0-norm minimization problem.

The major contributions of this paper are concluded as follows. A visual tracking method is proposed based on the primal l_0-norm minimization which can control the number of non-zero coefficients of all samples globally in sparse representation. We propose a multi-task version of l_0 sparse coding algorithm and present an effective algorithm to solve it. Multiple l_0-norm minimization problems could be solved by solving only a single optimization problem. It makes the procedure of l_0 sparse coding more computational efficiency. Experiments show that it can handle the difficulties during tracking such as partial occlusion and illumination changes.

2. l_0 sparse coding

In most sparse coding based object tracking algorithms, the optimization problem is often formulated as an l_1 or l_2 minimization problem. In this paper, we aim at directly addressing the primal l_0 sparse coding problem with l_0 gradient minimization [32,34], and introduce a novel iterative method which can achieve optimization solution for each iterative step to solve it approximately. We further reformulate multiple l_0 sparse coding problem into a single optimization problem, and present the multi-task l_0 sparse coding algorithm.

2.1. The primal L_0 sparse coding problem

Let $\mathbf{D}$ be a dictionary and denote $\mathbf{y}$ as a candidate to be encoded. To encode $\mathbf{y}$, the code $\boldsymbol{\alpha}$ can be calculated by using l_0-norm minimization

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2+\lambda\|\boldsymbol{\alpha}\|_0, \tag{1}$$

where λ is a constant that controls reconstruction error and sparsity.

The l_0-norm minimization problem is NP-hard. Now we introduce a novel method to solve the l_0-norm minimization problem approximately in [32]. Eq. (1) can be approximated as

$$\min_{\boldsymbol{\alpha},\mathbf{h}} \|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2+\lambda C(\mathbf{h})+\nu\|\boldsymbol{\alpha}-\mathbf{h}\|_2^2, \tag{2}$$

where $C(\mathbf{h})=\{q||h_q|\neq 0\}$ where h_q is the q-th element of $\mathbf{h}$ and ν is an automatically adapting parameter to control the similarity between $\mathbf{h}$ and $\boldsymbol{\alpha}$. Eq. (2) is equivalent to Eq. (1) when ν tends to infinity. This optimization problem is not jointly convex over $\mathbf{h}$ and $\boldsymbol{\alpha}$. Nevertheless, it can be solved by a way of alternatively minimizing $\mathbf{h}$ and $\boldsymbol{\alpha}$. We can optimize one parameter by giving another one and alternate between these two.

Given $\mathbf{h}$, to minimize the object function in Eq. (2) is equivalent to minimizing the following object function:

$$f(\boldsymbol{\alpha})=\|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2+\nu\|\boldsymbol{\alpha}-\mathbf{h}\|_2^2. \tag{3}$$

The optimization function $f(\boldsymbol{\alpha})$ is convex and has a global minimum. The solution can be obtained by gradient decent and calculated as

$$\boldsymbol{\alpha}=(\mathbf{D}^T\mathbf{D}+\nu\mathbf{I})^{-1}(\mathbf{D}^T\mathbf{y}+\nu\mathbf{h}), \tag{4}$$

where $\mathbf{I}$ is an identity matrix.

Given $\boldsymbol{\alpha}$, Eq. (2) is equivalent to minimizing

$$f(\mathbf{h})=\sum_i\left((\alpha_i-h_i)^2+\frac{\lambda}{\nu}P(h_i)\right), \tag{5}$$