Signal Processing: Image Communication 49 (2016) 17–24

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Visual tracking via adaptive multi-task feature learning with calibration and identification

Pengguang Chen[a,*], Xingming Zhang[b], Aihua Mao[b], Jianbin Xiong[a]

[a] School of Computer and Electronics Information, Guangdong University of Petrochemical Technology, Maoming 525000, China
[b] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

ARTICLE INFO

Keywords:
Multi-task feature learning
Object tracking
Feature selection
Sparse representation

ABSTRACT

Recently multi-task feature learning has become a widely applied approach for visual tracking, since it is benefited from the shared features across tasks. However, selecting features appropriately from multiple tasks is still a challenging problem due to the complex variation of the appearance of moving objects, which influences not only the features of single task but also the relationships between the features of multiple tasks. To address this problem, this paper presents a novel sparse learning model for selecting multi-task features adaptively. Compared to the existing multi-task models, the proposed model is capable of both calibrating the loss function according to the noise level of a task to keep its specific features, and identifying the relevant and irrelevant (outlier) tasks simultaneously by decomposing the regularized matrix into two specified structures. The proposed model allows to preserve specific features of individual tasks via calibration and to exploit sparse pattern over the relevant task via identification. Empirical evaluations demonstrate that the proposed method has better performance than a number of the state-of-the-art trackers on available public image sequences.

1. Introduction

Modeling an object's appearance effectively is one of the key issues for visual tracking [1]. To deal with the changes of objects' appearance influenced by illumination, shape deformation and occlusion, selecting shared features form a group of tasks, which can compensate the features of individual task [2–4], has more advantages than the single task method. Meanwhile, thanks to the stable recovery and sufficient dimensionality reduction of sparse signals [5], the feature selection from multiple tasks combining with sparse representation has robust performance in feature learning. Recently, the multi-task feature learning model, namely learning and selecting a set of shared sparse features from multiple tasks together, has been widely applied in visual tracking and is regarded as an effective method.

The existing multi-task feature learning algorithms proposed many strategies in improving the two components of the formulate, i.e., the loss function and regularizer. The former component is closely associated with the specific features of a task, while the later one is greatly dependant on the correlation between multiple tasks.

As to the loss function, it servers as quantifying the performance of a task corresponding to a linear model since multi-task feature learning model is based on a group of linear models. Usually, the optimization of non-convex function using convex relaxation cannot guarantee the solution to converge to an optimal solution [6], while the optimization of convex function allows a global optimal solution [7]. Thus the convex loss function has much superiority in achieving global solution during optimization. The typical convex loss function has a variety of forms, including hinge, least square and logistic functions. When compared with smooth loss function, the non-smooth loss function, such as $\ell_{1,p}$-norm [8] and group square-root lasso [9], usually obtains more robustness in complex situations. Moreover, loss function uses non-smooth convex $\ell_{1,2}$-norm instead of the most common smooth ℓ_2-norm, which helps to calibrate individual tasks and decrease their noise levels [2,10].

As to the regularizer, it reflects the correlation of different tasks. ℓ_p-norm with variable p is the most important index for the correlation, where increasing p indicates the degree of covariate sharing changing from none (p=1) [11] to full sharing ($p = \infty$) [12]. With the assumption of a strong constrain that all sparse patterns corresponding to individual tasks are strictly relate to sparse patterns with specific form, ℓ_2-norm regularization is employed in the early work [13]. As $\ell_{1,2}$-norm regularization encourages to pursue a similar sparse pattern among related tasks [7], it becomes a popular method in regularization. Recently, hybrid methods consisting of $\ell_{1,2}$-norm and variants of $\ell_{1,2}$-norm, such as group lasso, group square lasso and group square-root lasso, have been developed by many work. Moreover, by decom-

* Corresponding author.
E-mail addresses: chen.pg@mail.scut.edu.cn (P. Chen), cszxm@scut.edu.cn (X. Zhang), ahmao@scut.edu.cn (A. Mao), liangqiong8362@qq.com (J. Xiong).

http://dx.doi.org/10.1016/j.image.2016.09.009
Received 26 October 2015; Received in revised form 28 September 2016; Accepted 29 September 2016
Available online 03 October 2016

posing the weight matrix of the $\ell_{1,2}$-norm into two components, it is able to identify the outlier tasks by this kind of methods [4,14].

However, selecting feature from multiple tasks appropriately is still a challenging problem because of the sophisticated variations of appearance of the tracked objects [15]. On one hand, the variations of the essential content, structure and noise level of individual tasks usually lead to great change over their corresponding sparse patterns. On the other hand, changes of the tasks' relationships lead to the assumption that all the task share a common sparse structure not valid again. In practice, the common shared features of multiple tasks importantly depends both on the specific features of individual task and the correlation of multiple tasks. Although many multi-task feature learning methods have significantly extend the loss function and regularizer to generalized forms for robust performance, most of them just concern a single aspect of the task features. Due to that, they still have limitation in achieving the common features, especially in the cases of complex multiple tasks.

This paper aims at developing a new sparse learning model for adaptive selection of multi-task features for visual tracking. The proposed model focuses on both of the two aspects namely, the specific features of individual tasks and the correlations of multiple tasks, and thus can robustly select the shared features between multiple tasks. It calibrates the loss function according to the noise level of individual tasks for preserving their specific features, and also decomposes the regularized matrix into two given structures simultaneously to identify the relevant and outlier tasks for extracting the shared features across tasks. This model extends the traditional multi-task learning model to be capable of both calibration and identification, and meantime makes an integration of them to have more better performance in visual tracking.

The contributions of this paper is mainly in the following aspects:

- Presenting a new sparse learning model, which juxtaposes the specific features of individual tasks and the correlations of multiple tasks for adaptive selection of multi-task features. The proposed model has the advantages derived from both the calibrated model and the identified model.
- Developing an efficient optimization algorithm, which uses the dual smoothing technique on loss function and then solves the smoothed minimization problem via proximal gradient algorithm.
- Extracting shared features among multiple tasks by decreasing the noise level of tasks and discovering outlier tasks from them, which is effective to improve the tracking accuracy.

2. Related work

Sparse representation of the features can efficiently preserve the essential content of the moving object's appearance [16,17]. It is beneficial to use sparse representation to model the object's appearance for improving the robustness of approximation matching, as reported in [18,19]. Subsequently, lots of techniques are developed for improving the tracking accuracy, such as sparsity-based generative and discriminative models [20,21], local sparse appearance [22] and different feature descriptors [23,24]. Several work [5,25] also offer efforts in improving computation efficiency. All these sparse algorithms are based on a single task. However, selecting shared features form multiple tasks has the advantage to improve performance. Recently, multi-task feature learning model has been widely applied.

There are namely two progresses in multi-task feature learning model. One is to develop effective penalty term over the loss function. Using convex loss to replace non-convex loss become a robust approach for global optimization [7]. Specially, many algorithms employ smooth convex ℓ_2-norm as the loss function, such as [3,14]. For improving robustness, Nie et al. [8] and Bunea et al. [9] exploited non-smooth loss to replace smooth loss with $\ell_{1,p}$-norm and group square-root lasso, respectively. Furthermore, to reduce noise corruption and improve computation performance, non-smooth convex $\ell_{1,2}$-norm loss was applied in [2,10]. Another improvement comes from the contribution of the regularizer and its generalization capability. To improve the computation performance while reduce the computation complexity, [12,13] used ℓ_2-norm to replace ℓ_1-norm regularizer. With the generalization of ℓ_2-norm regularizer, $\ell_{1,2}$-norm became a popular method [7]. Moreover, algorithms proposed in [3,14] identified the relevant and outlier tasks by decomposing $\ell_{1,2}$-norm regularized matrix into two structures.

3. Adaptive multi-task feature learning for tracking

In this section, we present the details of the proposed tracking method which extracts features among multiple tasks adaptively. It consists of three parts: the introduction of multi-task tracking model, the model extension with calibration and identification, and the optimization procedure of the proposed algorithm.

3.1. Sparse-based multi-task tracking

In the sparse-based multi-task tracking framework, it is given a supervised learning set of K tasks associated with the dataset $\{(X^k, y^k)\}_{k=1}^K$, where $X^k \in \mathbb{R}^{n_k \times d}$ and $y^k \in \mathbb{R}^{n_k}$ denote the template matrix and the responding vector for the k-th task respectively; n_k is the number of samples; and d is the data dimensionality. When tracking target vector in the current image sequence, the target y^k can be robustly represented in a linear combination with template X^k and noise vector ε_k [3,26,27] as

$$y^k = X^k w^k + \varepsilon_k, \quad k = 1, \ldots, K, \tag{1}$$

where $w^k \in \mathbb{R}^d$ is the weight vector for the k-th task. Denote $W = [w^1, \ldots w^K] \in \mathbb{R}^{d \times K}$ as the weight matrix, and noise vector $\varepsilon_k \in \mathbb{R}^{n_k}$, $k = 1, \ldots, K$. Such condition induces to a minimization problem

$$\underset{W}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} (\phi(\varepsilon_k) l(y^k - X^k w^k)) + \lambda \Omega(W) \right\}, \tag{2}$$

where $\phi(\cdot)$ denotes a weighted function associated with the current task's noise; $l(\cdot)$ denotes a loss function; $\Omega(\cdot)$ is a regularization term, and λ is the nonnegative regularization parameter. Due to that, the tracking feature is dynamically preserved and represented via the weight matrix W.

Specifically, for feature selection in the problem (2), the traditional multi-task feature learning model with respect to W is usually formulated as:

$$\underset{W}{\operatorname{argmin}} \left\{ \sum \| y^k - X^k w^k \|_2^2 + \lambda \| W \|_{1,2} \right\}, \tag{3}$$

where $\|\cdot\|_2$ denotes the Euclidean norm, i.e., $\| X \|_2 = (\sum_i x_i^2)^{1/2}$, corresponding to the loss function $l(\cdot)$; and $\|\cdot\|_{1,2}$ denotes the $\ell_{1,2}$-norm of a matrix, i.e., $\| X \|_{1,2} = \sum_i (\sum_j x_{ij}^2)^{1/2}$, corresponding to the regularization term $\Omega(\cdot)$. Such model ignores the effect of noise, and exploits the shared sparsity directly over all the tasks.

In tracking objects, some task features may be sensitive to noise, and also be unreliable under variations cased by background, occlusion and deformation, especially in long sequences. That leads to that the traditional model may not to be valid for tracking. To drive an adaptive feature selecting model over multiple tasks, we make two improvements on the loss function and regularizer for eliminating noise and discovering unreliable task features, respectively, see details in Section 3.2.

3.2. Multi-task learning model with calibration and identification

In this paper, we assume that two conditions are given: (a) each noise vector is independent and uncorrelated with other tasks, i.e.,