Neurocomputing 166 (2015) 210–228

Contents lists available at ScienceDirect

Neurocomputing

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

Robust visual tracking based on incremental discriminative projective non-negative matrix factorization

Jianfang Dou [a,*], Jianxun Li [b], Qin Qin [b], Zimei Tu [b]

[a] Department of Automation and Mechanical and Electrical engineering, School of Intelligent Manufacturing and Control Engineering, Shanghai Second Polytechnic University, Shanghai 201209, PR China
[b] Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, PR China

A R T I C L E I N F O

Article history:
Received 8 November 2014
Received in revised form
30 January 2015
Accepted 16 March 2015
Communicated by: Yue Gao
Available online 22 April 2015

Keywords:
Visual tracking
Non-negative matrix factorization
Speeded up robust features
Delaunay triangulation
Template matching

A B S T R A C T

Visual tracking usually requires an object appearance model that is robust to changing illumination, partial occlusion, large pose and other factors encountered in video. Most existed visual tracking algorithms tend to drift away from targets and even fail in tracking them in presence of significant variation of the object appearance model or challenging situations. To address this issue, we propose a robust tracking algorithm based on discriminative projective non-negative matrix factorization and a robust inter-frame matching schema. The models of target and background are presented by the basis matrices of non-negative matrix factorization. In order to adapt the basis matrices to the variation of foreground and background during tracking, an incremental learning method is employed to update the basis matrices. A robust inter-frame matching by bidirectional method and Delaunay triangulation is adopted to improve the proposal distribution of particle filter, thus enhancing the performance of tracking. Template matching is used to correct the drift of the target if the result of discriminative part is unreliable. The proposed method is embedded into a Bayesian inference framework for visual tracking. Experiments on some publicly available benchmarks of video sequences demonstrate the effectiveness and robustness of our approach.

1. Introduction

Visual tracking is one of the most important components in a wide range of applications in computer vision, such as surveillance, human computer interaction, robot navigation and so on [1,2]. The purpose of visual tracking is to estimate the state of the tracked target in a video. It is usually formulated as a search task where an appearance model is first used to represent the target and then a search strategy is utilized to infer the state of the target in current frame. Therefore, how to effectively model the appearance of the target and how to accurately infer the state from all candidates are two key steps for a successful tracking system. Although a variety of tracking algorithms have been proposed in the last decades, visual tracking still cannot meet the requirements of practical applications. The main difficulty of visual tracking is to design a powerful appearance model which should not only discriminate the target from its surrounding background but also be robust to its appearance variations. For the former issue, some promising progresses have been achieved recently by considering visual tracking as a two-class classification or detection problem. Many elegant features in the field of pattern recognition can be used to discriminate the target from its background. However, the latter is very difficult to achieve since there are a large number of unpredictable appearance variations over time such as pose changes, shape deformation, illumination changes, partial occlusion and so on.

A variety of online tracking algorithms have been proposed to overcome these difficulties during tracking. These methods can be formulated in two different ways: generative model and discriminative model. To deal with the challenges mentioned above, the state-of-the-art algorithms focus on robust object representation schemes with generative appearance models and sophisticated classifiers.

Generative methods represent objects with models that have minimum reconstruction errors, and then the tracking is expressed as finding the most similar candidate to the target. Therefore, generative trackers only aim at matching the target appearance. Some generative tracking methods include mean shift tracker [3,4], eigen tracker [5], incremental tracker [6], covariance tracker [7], covariance tensor tracker [8], Frag Tracker [9],VTD tracker [10] and l1tracker [11]. Ross et al. [6] present the tracking method that trains a low-dimensional subspace representation, and fits online changes in the target appearance. In Ref. [11], a target candidate is sparsely represented as a linear combination of target templates and trivial templates that only have

* Corresponding author.
E-mail address: specialdays_2010@163.com (J. Dou).

http://dx.doi.org/10.1016/j.neucom.2015.03.076
0925-2312/

one nonzero element in each of them. The sparse representation problem is solved through a L1 minimization problem to solve the model tracking problem. Nie et al. [12] address single/cross-camera multiple-person tracking by graph matching to deal with main difficulties caused by the existence of occlusion in single-camera scenario and the occurrence of transition in cross-camera scenario. In recent years, there has also been the approach to use multiple tracker methods that work separately and combine the outputs to a final more robust tracking result [13]. Wu et al. [14] built a large benchmark on visual tracking and evaluated thoroughly the performances of over 29 trackers. Discriminative trackers aim to find a decision boundary that can best separate the target from the background, such as [15–20]. Grabner et al. [16] propose a novel on-line AdaBoost feature selection for tracking, ie OAB Tracker. Saffari et al. [17] proposed a visual tracking algorithm based on Online Random Forests. Avidan et al. [19] form a feature vector by every pixel in the reference image and an adaptive ensemble of classifiers is trained to separate the object from the background. Collins and Liu [18] build a confidence map by searching the most discriminative RGB color combination in each frame. Babenko et al. [15] adopt online multiple instances learning to be robust to occlusions and other image corruptions. Such a way refers to treat object tracking as a binary classification problem to distinguish between the positive sample (target) and negative samples (background). For using the background information, these methods demonstrate strong robustness to avoid distracters in the background. Li et al. [21] assumed that if the object's appearance or background changes drastically or continuously, which causes the underlying data distribution to keep changing, and proposed a novel and interesting semi-supervised CovBoost method which is utilizing the information provided by the three kinds of samples effectively when training the best strong classifier for tracking. Zhong et al. [22] perform visual tracking via weakly supervised learning from multiple imperfect oracles to address difficult object appearance update and drift problems.

Non-negative matrix factorization (NMF, [23]) decomposes a non-negative data matrix X into the product of two lower-rank non-negative factor matrices, i.e., $X \approx WH$. Due to the non-negativity constraints on both factor matrices W and H, NMF learns parts-based representation and brings much attention in practical tasks such as image processing [24] and data mining.

In this paper, we propose a robust online tracking algorithm based on discriminative projective non-negative matrix factorization and a robust inter-frame matching schema. The models of target and background are presented by the basis matrices of non-negative matrix factorization. In order to adapt the basis matrices to the variation of foreground and background during tracking, an incremental learning is employed to update the basis matrices. A robust inter-frame matching by the bidirectional method [20] and Delaunay triangulation [26,27] is adopted to improve the proposal distribution of particle filter, thus enhancing the performance of tracking. Template matching is used to correct the drift of the target if the result of discriminative part is unreliable. The proposed method is embedded into a Bayesian inference framework for visual tracking. Experimental results on several challenging video sequences demonstrate the effectiveness and robustness of our approach.

The contributions of this paper are as follows.

1) Integrating inter-frame matching into the framework of visual tracking, the two parts can complement to another, thus improving the performance of tracking;
2) A Robust inter-frame matching based on Bidirectional method and Delaunay Triangulation is adopted to enhance the robustness and accuracy of matching;
3) An incremental discriminative projective non-negative matrix factorization learning method is employed to update the target and background basis matrices, and can adapt to the variation of foreground and background during tracking;
4) The updated target and background basis matrices can eliminate the background matches and improve the transition model of target.

The remainder of this paper is organized as follows: in Section 2 we summarize the previous works most related to our work. The proposed robust tracking method IDNMF_VT (Incremental Discriminative Non-Negative Matrix Factorization For Visual Tracking) is described in Section 3, respectively. Experiments and results are provided and analyzed in Section 4. Finally, our work is summarized and conclusions are drawn in Section 5.

2. Related work

Zafeiriou et al. [28] proposed Discriminative NMF (DNMF) by incorporating Fisher's criterion to NMF. Yuan et al. [29] proposed projective NMF (PNMF) based on the linearization method. In particular, PNMF learns non-negative basis of the lower dimensional subspace and considers its transpose as the projection matrix, i.e., $X \approx WW^TX$. Since the learned projection matrix is non-negative, PNMF obtains non-negative coefficient for any new coming example because multiplication of non-negative matrix and non-negative vector produces non-negative vector. In addition, since PNMF implicitly induces $WW^T \approx I$, rows of W are approximately orthogonal. Moreover, since W is non-negative, the orthogonality implies that each column of W contains few nonzero entries. Therefore, PNMF implicitly learns parts-based representation. Bucak et al. [30] proposed an incremental non-negative matrix factorization algorithm (INMF) method, which can update its factors without much effort.

While referring to the image matching, one of the key stages is to guarantee that all found matches correct. The Random Sample Consensus (RANSAC) paradigm proposed by Fischler and Bolles [31] detects outlying data by first randomly selecting samples of the minimum number of data items required to estimate a given entity and then looking for consensus of the estimates among the samples. It can be seen that RANSAC algorithm selects samples randomly. The random sampling has some disadvantages: it increases the number of iterations. If there are a lot of data to be evaluated, the time needed during the calculation will increase obviously. The worst situation is that all components will be chosen and calculated. Above all, there are three problems for RANSAC: Firstly, there is no upper bound on the time it takes to compute the transformation model parameters; secondly, the number of iterations N computed is limited, so that the solution obtained may not be optimal, and it may not even be one that fits the data in a good way; last but not the least, it requires us to set specific threshold for certain problems. Delaunay triangulation [32] pervade computer vision. They not only provide a convenient and robust neighborhood representation for Voronoi tessellations of the image plane, but also provide a powerful geometric representation for volumetric information. Based on the uniqueness of Delaunay triangulation and the similarity of triangulations for the same scene in different images. For inter-frame matching, we extract Surf features and refine matches by bidirectional algorithm and Delaunay Triangulation.

Over the last decade, tracking methods using particle filters (also known as condensation or sequential Monte Carlo models) have demonstrated a noteworthy success in visual object tracking [34]. The popularity of these methods stems from their generality, flexibility, and simple implementation. Increasing the number of particles sampled at each frame tends to improve tracking performance, accompanied by a linear increase in computational complexity. As a result, researchers have devised algorithms to deal