Signal Processing: Image Communication 68 (2018) 13–25

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Visual tracking using Locality-constrained Linear Coding and saliency map for visible light and infrared image sequences

Meng Ding [a,b,*], Yao Yuheng [c], Li Wei [d], Yunfeng Cao [e]

[a] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Street, Nanjing, PR China
[b] Science and Technology on Electro-optic Control Laboratory, 25 Kaixuan street, Luoyang, PR China
[c] Viterbi School of Engineering, University of Southern California, 649 W 34th street, Los Angeles, USA
[d] Jin Cheng College, Nanjing University of Aeronautics and Astronautics, 88 Hangjin Street, Nanjing, PR China
[e] College of Astronautics, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing, PR China

ARTICLE INFO

Keywords:
Visual tracking
Visible light image
Infrared image
Locality-constrained Linear Coding
Spectral residual saliency
Particle filtering

ABSTRACT

As a development of sparse coding, while retaining the advantage of sparse coding in classification, Locality-constrained Linear Coding(LLC) greatly improves the time efficiency of appearance modeling. However, in order to further promote the performance of real-time and develop a tracking algorithm that can be applied to both visible light images and infrared images, this paper proposes a tracking algorithm using LLC and saliency map under the framework of particle filtering. It is universally acknowledged that number of particles determines the accuracy of tracker under the framework of particle filtering. Unfortunately, the increase in the number of particles leads to the augment of computational burden. Therefore, the basic idea of the proposed algorithm is to reduce the computational number of observation vectors while keeping the effective number of particles and achieve the goal of strengthening the real-time performance of tracker. The proposed algorithm firstly uses spectral residual to obtain a saliency map of the current frame and then computes the saliency score of each particle. Secondly, several particles are eliminated directly according to the difference between the saliency score of the particle in the current frame and the target score in the previous frame. Thirdly, LLC is used to compute the observation vector for the rest particles and complete tracking tasks. Both quantitative and qualitative experimental results demonstrate that the proposed algorithm performs favorably against the nine state-of-the-art trackers on twelve challenging test sequences including six visible light sequences and six infrared sequences. In addition, related experimental results reveal that the proposed algorithm decreases the computational complexity and has the better tracking performance compared with the tracker just using LLC in the framework of particle filtering.

1. Introduction

As an important and active research topic in the field of computer vision and pattern recognition, visual tracking can be applied in a wide range of domains [1,2]. With the development of sensor technology, it is possible to achieve robust target tracking under different visibility conditions. As an example, infrared sensors can offer clear infrared images in which the temperature and radiated heat determine the target intensity. Therefore, infrared images are not sensitive to illumination conditions and infrared sensors can be used in low visibility conditions, such as night, smoke, fog and haze. For a vision-based intelligent system, it is necessary to develop a tracking algorithm that can be used for both visible light images under high visibility conditions and infrared images under low visibility conditions.

Generally, visual tracking is the process that predicts the position of an object from an image sequence and can be formulated as a search problem that aims at finding the candidate region most matching to a target template as the tracking result. Target template is represented by an appearance model which evaluates the likelihood of a candidate region being the target [3]. The original target template comes from the target appearance in the first frame of the image sequence and is updated online once the tracking result is available in the following frames.

An excellent visual tracking algorithm should be robust to the extrinsic environment variations such as occlusion and illumination changes, and adaptive to the intrinsic appearance variations such as the target pose change and shape deformation. Traditionally, visual tracking algorithms can be divided into two categories, one is discriminative method and the other is generative method.

* Corresponding author at: College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Street, Nanjing, PR China.
E-mail address: nuaa_dm@nuaa.edu.cn (M. Ding).

https://doi.org/10.1016/j.image.2018.06.019
Received 19 March 2018; Received in revised form 25 June 2018; Accepted 28 June 2018
Available online 5 July 2018
0923-5965/

The discriminative methods usually formulate the tracking problem as a binary classification task to separate the target from the background. Unlike the general image classification problem, discriminative methods require updating constantly the classifier by online learning in the tracking process. Several of the classic online learning algorithms for tracking include online boosting (OAB) [4], online supervised boosting (SemiB) [5], online multiple instance learning (MIL) [6], online weighted multiple instance learning (WMIL) [7], kernelized correlation filters(KCF) and its improved algorithms [8–10]. Since classifier update is considered as a process of semi-supervised learning, the uncertainty of the labels corresponding to the new training samples from current tracking results may lead to the drifting problem. Generative methods typically learn an appearance model to represent the target and then formulate the visual tracking as an optimization problem that searches for the image patch with maximum likelihood or minimal error in the subsequent frames. Several of classic generative algorithms include adaptive structural local sparse appearance model (ASLA) [11], L1 tracker using accelerated proximal gradient approach (L1APG) [12], distribution fields for tracking (DFT) [13], etc. For generative methods, it is critical to construct a robust and effective appearance model that can adapt to various factors including the extrinsic and intrinsic appearance variations.

In general, tracking algorithm always needs to choose a search strategy. Many generative trackers use particle filter as the search strategy. Particle filter represents the target state by a set of weighted particles and the weight of each particle is re-assigned according to its contribution in the process of finding the target's location in the following frame image. The motion model and measurement data determine the position update of each particle. The number of particles is a considerable parameter for the robustness and real-time performance of the trackers. Generally, reducing the particle number can improve the real-time performance, but may result in the degradation of tracking accuracy.

Constructing an effective appearance model is of great importance in either discriminative or generative methods. In discriminative methods, the essence of appearance modeling is to extract features from the target and background to train a classifier. Taking KCF for example, the original feature extraction is gray value or HOG (Histogram of Oriented Gradient) feature. In the following improved versions, CNN (convolutional neural network) features are used in the framework of KCF [14,15]. However, due to the high dimensionality of deep features, the majority of tracking strategies cannot guarantee the real-time performance of the algorithm. In generative methods, the appearance model needs to be more robust to adapt to the environment changes and intrinsic appearance variations. Under the framework of particle filter, the purpose of appearance modeling is to measure the observation vector of particles.

Various feature representation schemes for tracking have been presented over several decades. In general, these schemes can be divided into three categories: primary feature, intermediate feature and advanced feature. As primary features, edge, contour, texture and color information are ubiquitous and used widely under the framework of filtering for tracking tasks [16–18]. Primary features can often offer a robust defense against noise, such as color histogram, but they may not perform well when illumination variations occur. In addition, due to the lack of color and texture information, using these features for infrared images may lead to that tracking tasks fail. In comparison with primary feature, intermediate feature has better abilities to distinguish the target from the background. For example, under the framework of KCF, the performance of tracking based on HOG feature is better clearly than based on gray value of the pixels [8], and fast Compressive Tracking (FCT) uses Harr feature to construct the appearance model [19].

In recent years, with the development of deep learning, advanced or deep feature has been widely used in different fields of computer vision. Unlike intermediate feature, which extracts the specific information of image patch, inspired by the working principle of human visual systems, the advanced feature is obtained by unsupervised learning. At present, the advanced features commonly used in tracking algorithms mainly include sparse coding and CNN-based feature. As a type of sparse coding, Locality-constrained Linear Coding(LLC) used in proposed algorithm is one of advanced features. Compared with the other two types of features, advanced features can extract richer and more comprehensive information of the target. Compared with CNN-based feature, the number of samples required for dictionary learning of LLC is small. This is very crucial for the infrared images because existing infrared image database is difficult to meet the needs of CNN training. Consequently, this paper chooses LLC as the basic approach of appearance modeling.

Furthermore, the current researches show that the trackers of combining particle filter and sparse coding can solve the problem of target offset [20]. However, the real-time of these trackers is a problem that has to be solved. The proposed algorithm will improve real-time performance in two aspects: The first is to utilize LLC to replace sparse coding for appearance modeling. Since LLC incorporates locality constraint to replace the sparsity constraint of sparse coding, the computational complexity of LLC is reduced greatly. Additionally, the sparsity is preserved because the locality of LLC is more essential than sparsity. The other aspect is that the proposed algorithm computes the saliency scores of all particles via spectral residual with low computation cost, and then selects some of these particles to extract observation vector by LLC according to the score of each particle. The computational cost is reduced by decreasing the number of the particles that need to measure the observation vectors by LLC.

In summary, the main contributions of this paper can be summarized as follows: (1) This paper proposes a tracking algorithm combining LLC and spectral residual. (2) Under the framework of particle filter, the new update mechanism of particle weight are designed according to the saliency score. (3) The proposed algorithm can be used for target tracking in both visible light and infrared images.

The rest of this paper is organized as follows: Section 2 discusses the related work and Section 3 presents the details of proposed tracking algorithm using LLC and spectral residual saliency under the framework of particle filtering. In Section 4, the experimental results and related comparisons on challenging sequences including visible light and infrared images are demonstrated. Finally, conclusion is summarized in Section 5.

2. Related work

Surveys of tracking methods can be found in many excellent reviews. Therefore, this section only discusses the methods closely related to our work including LLC and saliency analysis for tracking in details.

2.1. From sparse coding to LLC for tracking

For the past several years, sparse coding/representation and compressed sensing have attracted a great of attention in different fields of computer vision, including tracking issue. These trackers show state-of-the-art performance as reported in previous publications [21]. In general, sparse coding and compressed sensing can be used as a part of appearance modeling in either discriminative or generative trackers. As a classifier, sparse representation is usually used in discriminative trackers. This section divides the existing these trackers into three categories:

(1) *Sparse representation-based target searching* [22,23]: For the discriminative trackers, sparse representation classifier (SRC) can be used as a binary classifier. The dictionary of SRC contains two subspaces: one is target templates and the other is background regions. The sparse coefficients of candidate regions computed by *L0-norm* or *L1-norm* minimization determine the location of the target. Although SRC has some abilities to solve several challenges such as occlusion and target appearance variations during tracking, sparse representation-based target searching also has manifest shortages. Firstly, the dictionary