Neurocomputing 135 (2014) 118–129

Contents lists available at ScienceDirect

Neurocomputing

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

Robust visual tracking based on interactive multiple model particle filter by integrating multiple cues

Jianfang Dou, Jianxun Li

Department of Automation, Shanghai Jiao Tong University, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 2 August 2013
Received in revised form
15 December 2013
Accepted 15 December 2013
Communicated by Haowei Liu
Available online 28 January 2014

Keywords:
Visual tracking
Particle filter
Interactive multiple model
Corrected background weighted histogram
Completed local ternary patterns
Histogram of oriented gradients

ABSTRACT

Visual tracking can be formulated as a state estimation problem of target representation based on observations in image sequences. To investigate the integration of rough models from multiple cue and to explore computationally efficient algorithms, this paper formulates the problem of multiple cue integration and tracking to combine Interactive Multiple Model (IMM) with particle filter (IMM_PF). Interactive Multiple Model can estimate the multiple cue state of a dynamic system with several behavioral models that switch from one to another using model likelihoods and model transition probabilities. For the problem of visual tracking, the model of IMM is adopted to three target observation models: Corrected Background Weighted Histogram (CBWH), Completed Local Ternary Patterns (CLTP) and Histogram of Oriented Gradients (HOG). The probabilities of these models are corresponding to the weights of multiple cues. IMM_PF then dynamically adjusts the weights of different features. Compared with those state-of-the-art methods in the tracking literature, this algorithm can track the object accurately in conditions of rotation, abrupt shifts, as well as clutter and partial occlusions occurring to the tracking object with good robustness, as demonstrated by experimental results.

1. Introduction

Object tracking in a video sequence is an important problem in computer vision with applications in areas like video surveillance, vehicle navigation, perceptual user interface and augmented reality. Object surveillance may provide crucial information about the behavior, interaction, and relationship between objects of interest. Since automated video surveillance systems comprise modules from bottom to up processing, e.g., object detection, tracking, event analysis and classification, robustness and efficiency in each module are particularly important. However, the task of robust tracking is very challenging regarding illumination variation, background clutter, fast motion, occlusion, structural deformation, real-time restriction, etc. Over the years, a great number of tracking methods have been proposed to overcome these challenges. For a survey of these algorithms, we refer reader to [1,2].

Comaniciu et al. [3] proposed a background weighted histogram to decrease background interference in target representation. Ning et al. [4] demonstrated that the background weighted histogram based mean shift tracker was equivalent to the conventional mean shift tracking method [5], and then a corrected background weighted histogram is proposed to actually reduce the interference of background in target localization. Additionally, only utilizing color histogram to model the target object in visual tracking has the disadvantage that the spatial information of the target object is lost. Therefore, many other features, such as edge features [6], Local Binary Pattern (LBP) texture features [13], have been used in combination with color [7].

Local Binary Pattern (LBP), firstly introduced by Ojala et al. [9], is a simple yet efficient operator to describe local image pattern. It has a great success in computer vision and pattern recognition, such as face recognition [10], texture classification [11,12], unsupervised texture segmentation [13], dynamic texture recognition [14]. However, some issues still remain to be further investigated. Tan and Triggs [15] proposed Local Ternary Patterns (LTP) that quantizes the difference between a pixel and its neighbors into three levels, which is less sensitive to noise in near-uniform image regions such as cheeks and foreheads. Guo et al. [16] proposed Completed Local Binary Pattern (CLBP) where a local region is represented by its center pixel and a local difference sign-magnitude transform. CLBP is able to preserve an important kind of difference magnitude information for texture classification.

More recently, the Histogram of Oriented Gradients (HOG) descriptor and its variants have been used in human detection [17] and object class recognition [18], and have been shown to be very distinctive and robust, because they can model local object appearance and shape by the distribution of local intensity gradients or edge directions.

The Interactive Multiple Model (IMM) [19,20] estimator is a suboptimal hybrid filter that has been shown to be one of the most cost-effective hybrid state estimation schemes. The main feature

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

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.12.049

of this algorithm is its ability to estimate the state of a dynamic system with several behavior models which can "switch" from one to another. Particle filtering approaches for Markovian switching systems have been proposed in [21–23]. A major drawback of these methods is that there is no control over the number of particles in a model. In [24], a multiple model particle filter is proposed. However, there is no interaction between the models.

Combining particle filter with the IMM approach, an Interactive Multiple Model Particle Filtering (IMM_PF) method to improve the robustness of visual tracking based on multiple cue is presented. Similar to the standard IMM filters, the IMM_PF consists of four steps: interaction/mixing, filtering, model probability update, and state combination. In this paper, we investigate the integration of rough models from multiple cues and to explore computationally efficient algorithms, formulate the problem of multiple cue integration and tracking in particle filter framework based on combing Interactive Multiple Model (IMM) with particle filter. IMM can estimate the state of a dynamic system with several behavioral models that switch from one to another using model likelihoods and model transition probabilities. For the problem of visual tracking, the model of IMM is adopted to three observation models: Corrected Background-Weighted Histogram (CBWH), Completed Local Ternary Patterns (CLTP) and Histogram of Oriented Gradients (HOG). The models probabilities corresponding to the weights of multiple cues. IMM then dynamically adjusts the weights of different features.

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 IMM_PF_VT (Interactive Multiple Model Particle Filter 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

There is an extensive literature on object tracking. Due to space limitation, we only briefly review nominal tracking methods and those that are the most related to our own. For a more thorough survey of tracking methods, we refer readers to [1]. Object tracking methods can be categorized as either generative or discriminative. In generative tracking methods, a generative (possibly dynamic) appearance model is used to represent target observations. Here, the tracker searches for a potential target location that is most similar in appearance to the generative model. Popular generative trackers include mean shift tracker [3], L1 tracker [25] and incremental tracker [26]. For L1 minimization approaches, as reviewed in [27], there are several representative methods, such as Gradient Projection, Homotopy, Iterative Shrinkage-Thresholding, Proximal Gradient, Augmented Lagrange Multiplier (ALM), Least angle regression (LRS). Mei et al. [25] considered visual tracking as a two-subspace searching problem where the first subspace was spanned by a set of target templates and the second subspace was spanned by a set of trivial templates. The trivial templates are column vectors of an identity matrix, also called identity pixel basis. Discriminative trackers formulate the tracking problem as a binary classification problem. In this case, the tracker finds the target location that best separates the target from the background. Popular discriminative methods include on-line boosting [28], ensemble tracking [29], and online MIL tracking [30]. MIL can provide a powerful mechanism to deal with label ambiguities that are common in weakly annotated datasets. Andrews et al. [31] proposed two extensions of a support vector machine mi-SVM (maximizing instance margin) and MI-SVM (maximizing bag margin), that lead to mixed integer quadratic programs. MIL was linked to semi-supervised learning (SSL) in [32], by viewing MIL as a problem with unlabelled data but positive constraints. Xu et al. [33] proposed Multi-Instance Metric Learning (MIMEL) to learn an appropriate distance under the multi-instance setting.

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 improve the computational complexity of this tracking paradigm, e.g. the coarse-to-fine strategy in [35].

3. Proposed method

3.1. Interactive multiple model (IMM)

For maneuvering targets, it often occurs that the system switches between different models of operations. This requires the filer to account for changes of models. For this purpose, the multiple model (MM) approach and especially the interacting multiple model (IMM) [36] are commonly used. The MM approach has been proven to be an appropriate method for handling such nonlinear filtering problems. It has been shown that IMM estimator relatively performs much better than other MM methods [37] in case of maneuvering targets. In [38] a comparison between different types of IMM illustrates their concept of working and their performance at targets maneuvering for the three proposed scenarios. The IMM introduced good accuracy at minimum execution time. Munir and Atherton [39] proposed an adaptive interacting multiple model (AIMM) algorithm. The algorithm does not need to predefine the sub-models, but it needs to predefine and estimate the target acceleration according to the target motion characteristics; the correctness of the estimated acceleration significantly impacts the performance of maneuvering target tracking. However, since the models of IMM are the observation models and there is no need to estimate the target acceleration, the AIMM algorithm cannot meet the requirement of this paper.

The interacting multiple model, as one of the most efficient dynamic multiple model (MM) estimators, was proposed by Blom and Bar-Shalom. Different from many other methods which assume a particular moving pattern of the node, the IMM filter incorporates all the possible moving patterns of the node, by running a bank of filters parallel with each filter corresponding to one particular moving pattern. And the overall state estimate is a certain combination of these model-conditional estimates. A complete cycle of the IMM filter process consists of four essential operations, namely, mixing/interaction, filtering, model probability update, and combination. Each time it is running, it combines the appropriate models from its model database and chooses the combination of aircraft motion models that best fits the position data. After choosing the model, it is adapted to the aircraft dynamics in order to generate the best representation of the aircraft motion model [36]. The flow diagram of the IMM method is shown in Fig. 1where $M_{j,k}$ is model j at time $k,j=1,\ldots,r$; $\widehat{X}^{j}_{k-1|k-1}, \widehat{P}^{j}_{k-1|k-1}$ are the prior state estimate and its covariance for $M_{j,k}$; $\widehat{X}^{0j}_{k-1|k-1}, \widehat{P}^{0j}_{k-1|k-1}$ are the initial state estimate and its covariance for $M_{j,k}$; Z_k is the measurement at time k; $\widehat{X}^{j}_{k|k}, \widehat{P}^{j}_{k|k}$ are the state estimate and its covariance for $M_{j,k}$; Λ^{j}_{k} is the likelihood function for $M_{j,k}$; μ^{j}_{k} is the model probability for $M_{j,k}$; $\widehat{X}_{k|k}, \widehat{P}_{k|k}$ are the combined state estimate and its covariance.