Pattern Recognition Letters 33 (2012) 1075–1082

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Online multiple instance gradient feature selection for robust visual tracking

Yuan Xie [a], Yanyun Qu [b,*], Cuihua Li [b], Wensheng Zhang [a]

[a] State Key Lab. of Intelligent Control and Management of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] Video and Image Lab., Department of Computer Science, Xiamen University, Xiamen 361005, China

ARTICLE INFO

Article history:
Received 17 May 2011
Available online 10 February 2012
Communicated by A. Fernandez-Caballero

Keywords:
Gradient-based feature selection
HOG
Multiple Instance Learning
Online object tracking

ABSTRACT

In this paper, we focus on learning an adaptive appearance model robustly and effectively for object tracking. There are two important factors to affect object tracking, the one is how to represent the object using a discriminative appearance model, the other is how to update appearance model in an appropriate manner. In this paper, following the state-of-the-art tracking techniques which treat object tracking as a binary classification problem, we firstly employ a new gradient-based Histogram of Oriented Gradient (**HOG**) feature selection mechanism under Multiple Instance Learning (**MIL**) framework for constructing target appearance model, and then propose a novel optimization scheme to update such appearance model robustly. This is an unified framework that not only provides an efficient way of selecting the discriminative feature set which forms a powerful appearance model, but also updates appearance model in online **MIL** Boost manner which could achieve robust tracking overcoming the drifting problem. Experiments on several challenging video sequences demonstrate the effectiveness and robustness of our proposal.

1. Introduction

Visual tracking is a challenge problem in computer vision. It is well known that a good appearance model is very important for robust and efficient tracking. However, it is difficult to design a good appearance model because object target exhibits significant appearance change. Many tracking methods employ static appearance model such as works (Adam et al., 2006; Comaniciu et al., 2000; Isard and Maccormick, 2001; Avidan, 2001), these methods tend to fail when the appearance of the objects change significantly. As a result, there is the need for an adaptive appearance model to cope with appearance change during tracking. Therefore, in this paper we focus mainly on the following two points: (1) How to design an efficient appearance model. (2) How to update the appearance model in the online manner robustly.

With respect to modeling the object appearance, many works prefer to design adaptive appearance model using the current information both from the object and the background (Grabner et al., 2006; Grabner and Bischof, 2006; Jepson et al., 2003; Wang et al., 2005). Such way refers to treat object tracking as a binary classification problem which trains a model to separate the object from the background via a discriminative classifier.

Feature selection under boosting framework has been initially introduced by Tieu and Viola (2000) in the context of image retrieval. Then, Viola and Jones (2001) applied boosting feature selection to robust and fast object detection task, this was a seminal work that bridged the gap of weak learner design in boosting and feature selection step, paved the way of boosting in the area of computer vision, e.g., Opelt et al. (2004), Torralba et al. (2005), Yang et al. (2004). Those works take an exhaustive feature selection scheme over a very large hypothesis space. However, with the online constraint (Avidan, 2007; Grabner et al., 2006; Collins et al., 2005), the exhaustive feature selection over large feature space is strictly prohibited. To solve this problem, Grabner and Bischof (2006) proposed a novel feature selection method where a set of selectors were constructed to choose the feature by minimizing the training error from random guess feature pool.[1] But, it only picks up the most discriminative feature from such feature pool, if those features in the pool are less powerful, such scheme is far from efficiency. Liu and Yu (2007) propose a novel feature selection scheme for online boosting based on the gradient descent mechanism. The approach iteratively updates the features (Histogram of Oriented Gradient (Dalal and Triggs, 2005), but not limited to it) in a gradient descent manner. It seems that gradient feature selection is a much more efficient scheme of learning discriminative features.

The traditional tracking under boosting framework is a supervised learning method, therefore sampling the object and background is a critical step for updating the appearance model. Nevertheless, the inaccuracy in samples will degrade the

* Corresponding author. Tel.: +86 010 82614489; fax: +86 010 62545229.
E-mail addresses: yuan.xie@ia.ac.cn (Y. Xie), yyqu@xmu.edu.cn (Y. Qu), chli@xmu.edu.cn (C. Li), wensheng.zhang@ia.ac.cn (W. Zhang).

[1] The size of such random guess feature pool is relatively small comparing with the large feature space.

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2012.01.020

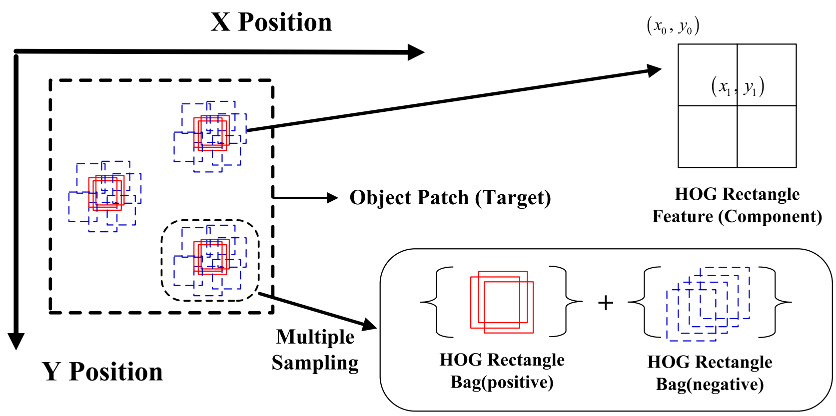

Fig. 1. Image representation and object appearance model.

appearance model and cause drifting. Thus, Viola et al. (2005) and Babenko et al. (2009) introduce the use of a Multiple Instance Learning (**MIL**) (Dietterich et al., 1997) for object detection and tracking. In fact, many applications of concept learning, unambiguously labeled positive and negative samples are not easily available. For example, in the context of object detection, a positive bag could contain a few possible bounding boxes around each labeled object without knowing which one is true "correct", therefore the ambiguity is passed to the **MIL** learning process. Then, **MIL** could handle such ambiguity by minimizing the negative log likelihood of training bags, so a more robust learner could be achieved. However, **MIL** based tracking (Babenko et al., 2009) still employs the exhaustive feature selection mechanism to form the adaptive appearance model which takes the negative influence on the power of tracking system.

This paper proposes a gradient-based feature selection unifying with online Multiple Instance Learning (**MIL**) approach for robust object tracking. The first contribution is the introduction of a novel discriminative object appearance model, such model consists of **HOG** rectangle features and their corresponding feature bags from positive and negative samples. The second contribution is the proposal of an *optimization scheme* updating appearance model robustly. It iteratively updates each feature using gradient descent and **MIL** approaches by maximizing the likelihood of training **HOG** feature bags. The proposed method not only provides an efficient way of building a discriminative appearance model, but also updates model robustly to drifting problem which traditional supervised learning unavoidable encounters. We present the empirical results of our method comparing with several state-of-the-art tracking algorithms on standard challenging video sequences, experimental results show that our method can lead to a more robust and stable tracker than state-of-the-art methods.

The rest of the paper is organized as follows: Section 2 gives an overview of the proposed tracking method. Then we give a short review of gradient descent feature selection and the online Multiple Instance Learning in Sections 3 and 4. Section 5 describes a novel online multiple instance gradient feature selection approach for object tracking. The experimental results and some discussions are reported on Section 6. Finally, we concludes the paper and outlines the direction of the future work on Section 7.

2. System overview

Usually, an object tracking system contains three components: image representation, object appearance model and motion model. We employ Histogram of Oriented Gradient (**HOG**) (Dalal and Triggs, 2005; Laptev, 2004) rectangle feature as the image representation to describe the components[2] of tracking object. Fig. 1 show the structure of **HOG** feature and we will discussed in more detail in Section 3.1. Our appearance model adopts the philosophy of representing object as an assembly of component (Fergus et al., 2005; Leibe et al., 2008; Kwon and Mu Lee, 2009). But unlike to such method, we use the bag of **HOG** rectangles to model a component, and apply boosting framework to combine the **HOG** components into a strong discriminant classifier, which is able to return $p(y=1|x)$ ($p(y|x)$ for short) where x is an image patch[3] and y is a binary variable indicating whether the x is the target. For motion model, supposed at time step $t-1$, our tracker maintains the object location l^*_{t-1}. Then, the simple distribution of target's location l^*_t at t step is:

$$p(l^*_t|l^*_{t-1}) \propto \begin{cases} 1 & \text{if } \|l^*_t - l^*_{t-1}\| < s, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where s is the radius of search region. The overview of our tracking system is summarized in **Procedure**. The workflow of our tracking system is similar to Algorithm 1 in the work (Babenko et al., 2009) with differences of central steps such as appearance model and its updating scheme (e.g., step 2 and 3). Supposed we are going to locate the target in the new frame at time step t. The method crops a set of image patch candidates within search region s, then computes their **HOG** components' feature vectors efficiently via Integral Histogram. The proposed method achieves the new location of target by means of **Online Gra-MIL** classifier $\boldsymbol{F}(x)$ (Section 5) trained in previous step, as can be shown from step 3 and 4 in **Procedure**. Accordingly, for each **HOG** component, we can crop the positive and negative **HOG** rectangle bags from positive patches bag $\boldsymbol{X}^{\gamma}$ and negative patches bag $\boldsymbol{X}^{\gamma,\beta}$ respectively at certain position and specified size[4] (step 5 in **Procedure**). Then each **HOG** component will be updated via a new optimization schema which employs gradient feature selection to maximize the likelihood of the current **HOG** feature bags. Consequently, the whole appearance model could be updated under the boosting framework, and the final strong classifier $\boldsymbol{F}(x)$ can be used to the next frame $t+1$ (see more details in Section 5).

[2] We call **HOG** rectangle features as components rather than parts, because they are not semantic parts of object like arms or legs of human.

[3] In fact, x is usually the representation of an image patch in feature space.

[4] The position and size are the variables which are the elements of parameter of **HOG** feature classifier, see more details in Section 3.