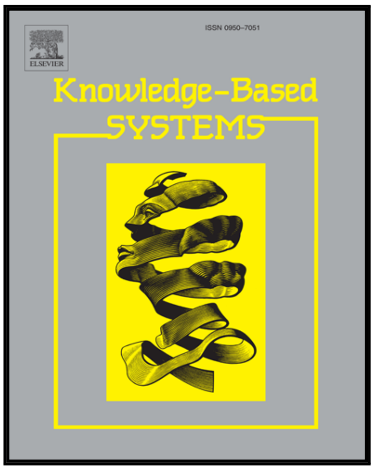
Accepted Manuscript



Object Tracking via a Cooperative Appearance Model

Fan Li , Shanshan Liu

PII: S0950-7051(17)30223-X
DOI: 10.1016/j.knosys.2017.05.012
Reference: KNOSYS 3910

To appear in: *Knowledge-Based Systems*

Received date: 10 January 2017
Revised date: 14 May 2017
Accepted date: 16 May 2017

Please cite this article as: Fan Li , Shanshan Liu , Object Tracking via a Cooperative Appearance Model, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.05.012

Object Tracking via a Cooperative Appearance Model

Fan Li, Shanshan Liu
School of Electronic and Information Engineering, Xi'an Jiaotong University 710049, Xi'an, China
lifan@mail.xjtu.edu.cn, xianghan@stu.xjtu.edu.cn

Abstract

Object tracking has been wildly used in security monitoring, traffic control, medical imaging and other fields. Conventional algorithms design local and holistic appearance model for effectively tracking. However, the performance of algorithms decreases in the complex scenes, including deformation and background cluster, partial or full occlusion and so on. In this paper, an object-tracking algorithm via a cooperative appearance model is proposed. Considering the visual characteristics of human eyes, we propose the impact region with different impact factors. The impact regions are defined as the regions which play different effects in making decision. The pixels in the regions with different distance from the target center will have different importance. We divide the impact regions into the significant impact region, the sub-impact region, and the non-impact region. The cooperative appearance model uses local collaborative representation to rectify holistic representation with impact regions. In local representation, positive and negative dictionary are derived from the candidates of video frames. The candidates are segmented into non-overlapping sub-blocks, and the sub-block responses of each candidate are obtained based on a collaborative dictionary. In holistic representation, the candidates are represented sparsely to obtain the total reconstruction error. The tracking result is decided by combing the sub-block responses in local representation and the reconstruction error in holistic representation with impact regions. The experimental results show that the proposed algorithm has performed well on deformation and illumination variation, partial or full occlusion, scale variation and background cluster compared with the state-of-the-art algorithms.

Keywords: object tracking; cooperative; local representation; holistic representation; impact region

1 Introduction

Object tracking is the determination of a target state, such as position, size or speed, in successive video frames. It is one of the key technologies in the field of computer vision, which is widely used in security monitoring, traffic control, medical imaging and other fields. It brings not only new hope to science and technology but also convenience to society and living. However, many challenges are faced in its implementation. These include determining how to model the target using a robust feature model under illumination variation (IV), deformation (DEF), scale variation (SV), in-plane rotation (IPR) and out-of-plane rotation (OPR), and how to search for the target quickly and efficiently from the candidate samples in the complex scenes with partial or full occlusion (OCC) and background clusters (BC). These problems will undoubtedly affect the accuracy of state estimation, resulting in tracking drift or even tracking failure.

To address these problems effectively, the existing tracking algorithms design different appearance models to determine the state of the object in continuous video frames (assuming the object state obeys the first-order Markov process [1]). The appearance model represents the similarity between the observations and the real states. It can be divided into generative and discriminative models. A generative model learns the distribution of the input data online and regards the conditional probability distribution as the model for describing the target. The model, combined with sparse representation, selects the sample with the lowest reconstruction error as the tracking result from the candidate samples. In the discriminative method, the inputs for online training are the data and labels, and the output is the set of class boundaries (decision functions), regarded as the discriminatory model. The basic strategy of this model is to treat the tracking problem as a binary problem, separating the target from the background.

Tracking based on the generative appearance model [2-5] usually uses a holistic template set and a sparse representation coding candidate sample. It makes full use of the holistic information of the foreground to represent the candidate samples. It selects the sample with the lowest reconstruction error as the current tracking result. Though tracking objects can be determined in many scenes, it ignores discriminative information for distinguishing foregrounds and backgrounds. In