Deep learning assisted robust visual tracking with adaptive particle filtering

Xiaoyan Qian, Lei Han Yuedong Wang, Meng Ding

PII: S0923-5965(17)30154-6
DOI: http://dx.doi.org/10.1016/j.image.2017.09.001
Reference: IMAGE 15274

To appear in: *Signal Processing: Image Communication*

Received date : 2 March 2017
Revised date : 5 September 2017
Accepted date : 5 September 2017

Please cite this article as: X. Qian, L.H.Y. Wang, M. Ding, Deep learning assisted robust visual tracking with adaptive particle filtering, *Signal Processing: Image Communication* (2017), http://dx.doi.org/10.1016/j.image.2017.09.001

Deep learning assisted robust visual tracking with adaptive particle filtering

Xiaoyan Qian, Lei Han Yuedong Wang, Meng Ding

College of Civil Aviation

Nanjing University of Aeronautics and Astronautics, 210016, Nanjing, China

Abstract

We propose a novel visual tracking algorithm based on the representations from a pre-trained Convolutional Neural Network (CNN). Our algorithm pre-trains a simplified CNN using a large set of videos with tracking ground truths to obtain a generic target representation. When tracking, Particle Filtering (PF) is combined to the fully-connected layer in the pre-trained CNN. Deep representations and hand-crafted features help to model tracking. To optimize the particles' distribution, the velocity and acceleration information aids to calculate dynamic model. Meanwhile, our algorithm updates the tracking model in a lazy manner to avoid shift and expensive computation. As compared to previous methods, our results demonstrate superior performances in existing tracking benchmarks.

Keywords: Visual tracking; Deep learning; Particle filter

1. Introduction

Visual tracking, i.e., tracking a specific target object in consecutive video frames to get its moving trajectory, is a fundamental problem in computer vision and has been actively studied for decades. A wide range of applications rely on robust visual tracking including security and surveillance, vehicle transportation, traffic monitoring and video compression [1].

Current visual tracking algorithms include generative and discriminative approaches. Generative methods establish appearance prior distribution and search for the target regions that fit the models best. Various generative appearance modeling includes template-models [2-4], incremental subspace learning [5], sparse representation [6,7] and back-propagation [8,9]. They can produce strong appearance features for the targets but lack of background information. So noises are easy to be brought into models once the occlusion appears, that will result in error and shift. In contrast, discriminative methods aim to build a model that distinguishes the target object from the background. These tracking algorithms