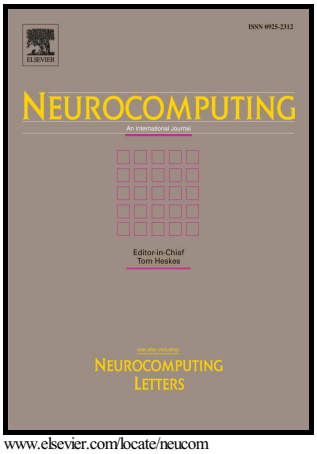

www.elsevier.com/locate/neucom

Dual-scale structural local sparse appearance model for robust object tracking

Zhiqiang Zhao, Ping Feng, Tianjiang Wang, Fang Liu, Caihong Yuan, Jingjuan Guo, Zhijian Zhao, Zongmin Cui

PII: S0925-2312(16)31048-7
DOI: http://dx.doi.org/10.1016/j.neucom.2016.09.031
Reference: NEUCOM17560

To appear in: *Neurocomputing*

Received date: 6 January 2016
Revised date: 24 July 2016
Accepted date: 13 September 2016

Cite this article as: Zhiqiang Zhao, Ping Feng, Tianjiang Wang, Fang Liu Caihong Yuan, Jingjuan Guo, Zhijian Zhao and Zongmin Cui, Dual-scale structural local sparse appearance model for robust object tracking *Neurocomputing,* http://dx.doi.org/10.1016/j.neucom.2016.09.031

Dual-scale Structural Local Sparse Appearance Model for Robust Object Tracking

Zhiqiang Zhao[a,b], Ping Feng[a], Tianjiang Wang[a,*], Fang Liu[a], Caihong Yuan[a,c], Jingjuan Guo[a,b], Zhijian Zhao[d], Zongmin Cui[b]

[a]School of Information Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[b]School of Information Science and Technology, University of Jiujiang, Jiujiang, Jiangxi 332005, China
[c]School of Computer and Information Engineering, Henan University, Kaifeng, HeNan 475004,China
[d]Business School of Hunan University, Changsha, HuNan 410006,China

Abstract

Recently, sparse representation has been applied in object tracking successfully. However, the existing sparse representation captures either the holistic features of the target or the local features of the target. In this paper, we propose a dual-scale structural local sparse appearance (DSLSA) model based on overlapped patches, which can capture the quasi-holistic features and the local features of the target simultaneously. This paper first proposes two-scales structural local sparse appearance models based on overlapped patches. The larger-scale model is used to capture the structural quasi-holistic feature of the target, and the smaller-scale model is used to capture the structural local features of the target. Then, we propose a new mechanism to associate these two scale models as a new dual-scale appearance model. Both qualitative and quantitative analyses on challenging benchmark image sequences indicate that the tracker with our DSLSA model performs favorably against several state-of-the-art trackers.

Keywords: Appearance model, visual tracking, sparse representation, dual scale

1. Introduction

Visual tracking [1] has been an important application field in computer vision, especially for the application of visual surveillance, traffic flow monitoring, robotics, and human-computer interaction. In the past decades, the technology of visual tracking has made great progress. However, visual tracking technology is still very challenging because there are uncertain factors in the tracking process, such as the change of target shape, illumination change, varying viewpoints, background clutter, partial occlusion etc.

Generally, a tracker includes three main parts: (1) a motion model, which produces candidate targets for the next frame to forecast the target, on the basis of the tracking result of the current frame; (2) an appearance model, which is used to model the shape of the target in the whole process of tracking; (3) an observation model, which defines a similarity measure between candidate targets and the target to discriminate the best candidate target. This paper focuses on the second part. An appearance model [2] generally includes two parts: visual representation and statistical modeling. The focus of visual representation is how to construct robust target descriptors using different visual features, while the focus of statistical modeling is to establish an effective mathematical model for the target by statistical learning.

Visual representation in visual tracking comprises holistic visual representations and local visual representations. The advantages of holistic visual representations are simplicity and efficiency. Raw pixel representation, histogram representation, covariance representation, and so on are classical holistic visual representations. Ross et al. [3] and Silveira and Mails [4] present raw pixel representation based on vectors. Other researches [5]-[7],[48,49] present the histogram representation as visual features of the target. The histogram representation can capture the holistic visual features of the target, but it is sensitive to the noise and the variety of illumination, as well as the shape of the target. A few other reports [8]-[10] present covariance representation as the visual features of the target. The advantage of covariance representation is that it can capture the intrinsic information of the target, but it is also sensitive to the noise and it loses the information on the shape and the location.

Holistic visual representation is sensitive to the holistic appearance variance, while local visual representation can capture more information on the local structural aspects of the target. Thus, local visual representation is more robust to variations in appearance of the target. Liu et al. [10] utilize a hierarchical part-template shape model to detect and segment humans. Liu et al. [11] use a particle-filtering (PF) model based on a template to track a target. Researchers of [10],[12] and [13] have also exploited a visual feature based on the superpixel for visual tracking. There are also papers [14]-[16] that present visual features based on Scale-invariant feature transform (SIFT). The advantage of these local visual representations is that

*Corresponding author
Email address: tianjiangwang2016@gmail.com (Tianjiang Wang)