Visual Tracking with Multilevel Sparse Representation and Metric Learning

Baifan Chen, School of Information Science and Engineering, Central South University, Changsha, China
Meng Peng, School of Computer and Communication, Hunan Institute of Engineering, Xiangtan, China
Lijue Liu, School of Information Science and Engineering, Central South University, Changsha, China
Tao Lu, Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China

ABSTRACT

Visual tracking arises in various real-world tasks where an object should be located in a video. Sparse representation can implement tracking problems by linearly representing object with a few templates. However, this approach has two main shortcomings. Namely, setting the templates updating frequency is difficult and meanwhile it is relatively weak in distinguishing the object from the background. For solving these problems, the author models a multilevel object template set that can be stratified by different updating time spans. The hierarchical structure and updating strategy promise the real-timeness, stability, and diversity of object template. Additionally, metric learning is combined to evaluate the object candidates and thereby improve the discriminative ability. Experiments on well-known visual tracking datasets demonstrate that the proposed method can track an object more robustly and accurately compared to the state-of-the-art approaches.

KEYWORDS
Metric Learning, Sparse Representation, Updating Strategy, Visual Tracking

INTRODUCTION

Visual tracking, also known as locating an object over time in a video or a visual sequence, has been widely used in multiple fields, such as intelligent monitoring, human-machine interaction, motion analysis, and vehicle navigation. Most of tracking methods are based on various appearance-learning models, and these models are always built from sample data online or offline to represent the tracked object. In general, appearance models consist of two main categories: generative model (Mei & Ling, 2011; Bao, Wu, Ling, & Ji, 2012; Ross, Lim, & Lin, 2008) and discriminative model (Babenko, Yang, & Belongie, 2011; Grabner, & Bischof, 2006; Avidan, 2007). To be specific, generative tracking methods generate the object appearance model by learning the sample’s distribution and searching for the most similar image regions. Discriminative methods consider tracking as a binary classification problem between the object and the background according to the decision boundary.

In recent years, visual tracking methods based on sparse representation have been applied successfully (Mei & Ling, 2011; Bao, Wu, Ling, & Ji, 2012). According to sparse representation theory, the object can be sparsely and linearly represented by a combination of templates. What is more, these methods can properly solve some classic problems of tracking, such as obscured objects, damaged images, the nonlinear object shape, and low-dimensional manifold structure. However, sparse trackers have two problems: (1) How often do the object templates need update? Indeed, the templates

DOI: 10.4018/JITR.2018040101

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
need update constantly due to frequent changing of the object appearances (Matthews, Ishikawa, & Baker, 2004). If the updating is too slow, the templates will not adapt to the morphological changes of the object, and meanwhile, the tracking accuracy will be low. If the updating is too fast, some parts of the background will be easily introduced into the templates, which will thereby result in serious drifting. (2) The sparse representation, which uses only a few observation samples, cannot accurately describe the object appearance. Therefore, sparse representation method is relatively weak in distinguishing the object from the background. Because the object template contains not only the object features but also background information, the complex background environment gives rise to unstable and poor controllable tracking results.

Templates are always updated by replacing (Mei & Ling, 2011; Bao, Wu, Ling, & Ji, 2012) and learning (Li, Shen, & Shi, 2011; Chen, Wang, Wang, Zhang, & Xu, 2011; Wang, Lu, & Yang, 2013; Zhang, & Li, 2011; Mairal, Bach, Ponce, & Sapiro, 2009). When the tracking result is different from the current template, the template will be replaced by the one with higher coefficient or higher weight, which means that it is closer to the appearance of the tracking result. In contrast to these replacing approaches, another way is using multiple subspaces learned to model the target appearance variations during the tracking to deal with abrupt appearance changes. It is always assumed that multiple linear models can approximate the appearance manifold of an object. However, since the appearance manifold is always nonlinear and complex, templates’ learning is easy to over-fit and cannot maintain local relationships of the manifolds.

In order to solve the above problems, this paper has proposed a multilevel hierarchical object template set as well as its updating strategy. Different layers in the template set store the object template with different updating frequency. Besides, the updating time span between the adjacent templates rises as their layer number increases. The strategy of templates updating follows the rule of low-to-high, inner-to-interlayer. Therefore, in a multilevel stratified object template set, the higher frequent updating at the lower level ensures the real-timeness, while the lower frequent updating at the higher level improves the stability and diversity. Furthermore, the metric learning theory is taken into account to compensate for weak discriminating ability of the sparse representation. Moreover, a visual tracking method by fusing multilevel sparse representation with metric learning is proposed. This tracking method is implemented in the particle filter framework, and the confidence value of the object candidate is calculated by combining multilevel sparse representation and metric learning. According to the experimental results, the method proposed in this paper has high robustness and accuracy.

MULTILEVEL SPARSE REPRESENTATION

Sparse Representation

This paper casts tracking as a sparse approximation problem (Mei & Ling, 2011). In a new image frame, the tracking object is regarded as the one with the smallest projection error from object candidates. Given that the object template set T includes L object templates, \( T = [t_1 \ t_2 \ \ldots \ t_L] \).

Each object template \( t \) refers to a 1D vector. Considering the effect of noise and partial occlusion, an object candidate \( y \) can be represented as:

\[
y = Ta + \varepsilon
\]  

(1)

where \( a \) denotes a coefficient vector, and \( \varepsilon \) denotes the error vector. The nonzero entries of \( \varepsilon \) indicate the pixels in \( y \) that are corrupted or occluded.

After this paper models error with trivial templates and adds non-negativity constraints with negative trivial templates, it models the object candidate as:
Library Networking of the Universidad de Oriente: A Case Study of Introduction of Information Technology
www.igi-global.com/article/library-networking-universidad-oriene/44598?camid=4v1a

Artificial Neural Networks in Financial Trading
www.igi-global.com/chapter/artificial-neural-networks-financial-trading/14230?camid=4v1a