On Fast Trackers that are Robust to Partial Occlusions

Lu Zhang Hamdi Dibeklioğlu Laurens van der Maaten
Computer Vision Lab, Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
{lu.zhang, h.dibeklioglu, l.j.p.vandermaaten}@tudelft.nl

Abstract

Model-free tracking aims identify the location of particular objects or object parts in each frame of a video based on a single positive example. In our work, we (1) develop online-learning algorithms for part-based models that facilitate the use of these models in model-free tracking in order to improve robustness to partial occlusions; and (2) derive a probabilistic bound that facilitates rapid pruning of candidate locations in many popular trackers. Together with other recent advances in object detection and tracking, we believe these developments will ultimately contribute to solving the long-term tracking problem.

1. Introduction

The goal of (long-term) tracking is to accurately identify the location of particular objects or object parts in every frame of an image sequence. To do so, a tracker requires an object appearance model that is invariant to all the appearance variations the object may exhibit whilst also capturing all information that is necessary to discriminate the object from all other objects in the image sequence (the "background"). In model-based trackers, such an appearance model is obtained by training on large annotated image collections such as ImageNet. The key disadvantage of such an approach is that it requires the collection and annotation of a large database of images that contain the target object, which is time-consuming and expensive. By contrast, model-free trackers aim to learn the appearance model based on a single annotation of the target object, thereby facilitating the rapid development of a new tracker. The model-free tracking scenario poses a difficult problem to the learning algorithm, namely, to learn how to discriminate the target object (whose appearance may strongly vary over time) in an online manner from a small set of observations with extremely limited availability of ground-truth. Moreover, it is essential that a tracker (model-free or model-based) runs in real-time, which limits the complexity of the algorithms that can be used.

In this work, we present two developments that aim to improve the quality and speed of modern model-free trackers. First, we present an online-learning approach for deformable template models based on pictorial structures, called structure-preserving object tracking [2, 3]. The use of pictorial-structures models allows us to learn more discriminative appearance models, even in very difficult tracking scenarios in which multiple, visually similar objects are present in the images. Second, we present a probabilistic bound that allows to speed up most state-of-the-art trackers by reducing the number of candidate locations for the target object that the tracker needs to inspect [1]. Both developments are briefly described below.

2. Structure-Preserving Object Tracking

A key problem of most appearance models in (model-free) tracking is that they are not robust to partial occlusions of the target object. We propose a new tracking algorithm, called structure-preserving object tracking (SPOT), that uses a part-based appearance model based on pictorial structures. The appearance model represents the target object with a "global" appearance model and a constellation of "part" appearance models that are spatially related. These spatial constraints are learned along with the object detectors using an online structured SVM algorithm. The key advantage of the use of the additional part models is that some parts of the target object can be appropriately identified even when the target object is partially occluded. The identified parts can then be used to infer the locations of the other parts and of the complete object. This reduces the tendency of existing trackers to drift under the influence of partial occlusions. Some of our experimental results are shown in Figure 1, which compares the performance of our SPOT tracker with two different types of spatial graph structures (star-SPOT and mst-SPOT) to that of an identical tracker without part models (no-SPOT) and two alternative trackers (OAB and TLD) in terms of average location error. The results presented in the figure highlight the potential of incorporating part models in the appearance models of trackers.

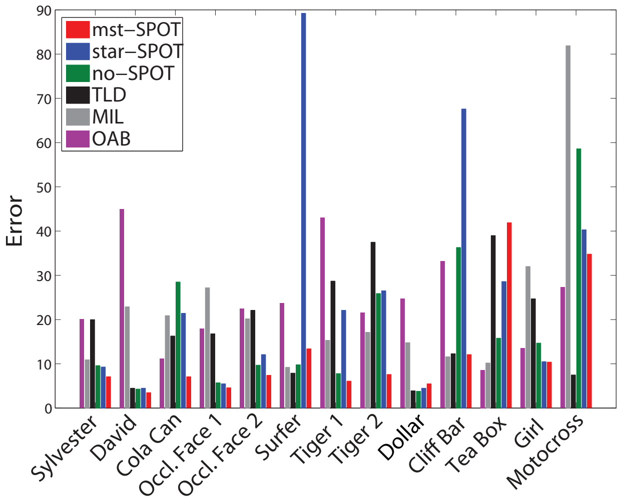

Figure 1. Average location error of five object trackers on thirteen test videos (lower is better). Figure best viewed in color.

3. Pruning Candidate Locations

To find the image location that best matches the object appearance model, most modern detectors and trackers use a form of sliding-window search or selective search: they compute a classifier score for a substantial number of locations and output the location with the highest score. We propose an algorithm to speed up such exhaustive searches by exploiting the fact (1) that classifier scores often decompose into a sum over features and (2) that feature values are generally bounded below and above. Specifically, these assumptions allow us to infer that the target object is not present at a location after considering only a small subset of features. For instance, if the classifier subscore of a particular location is a large negative number after considering 10% of the features whilst other locations have a very large positive subscore, this location may perhaps be discarded immediately. In fact, we can derive exact upper and lower bounds on the final classifier score for that location by exploiting the boundedness of the features, allowing to safely prune the location. We derive a probabilistic version of such bounds (based on a Chernoff bound) that allows the algorithm to discard many candidate locations early on whilst only allowing a very small probability of error. Some results of our experiments are shown in Figure 2, which shows the average location error of our tracker as a function of the percentage of features that it evaluates. The results in the figure reveal that our probabilistic bound may reduce the average number of inspected features by up to 90% whilst hardly affecting the accuracy of the tracker.

Acknowledgements

This work was supported by EU-FP7 Social Signal Processing (SSPNet) and by the China Scholarship Council.

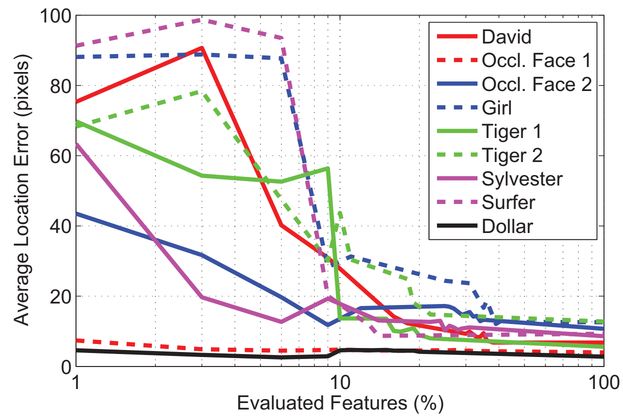

Figure 2. Average location error as a function of the percentage of features inspected by our tracker on nine test videos (lower is better). Figure best viewed in color.

References

[1] L. Zhang, H. Dibeklioglu, and L. van der Maaten. Speeding up tracking by ignoring features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[2] L. Zhang and L. van der Maaten. Structure preserving object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1838–1845, 2013.

[3] L. Zhang and L. van der Maaten. Preserving structure in model-free tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(4):756–769, 2014.