Li K, He FZ, Yu HP. Robust visual tracking based on convolutional features with illumination and occlusion handing. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 33(1): 223–236 Jan. 2018. DOI 10.1007/s11390-017-1764-5

Robust Visual Tracking Based on Convolutional Features with Illumination and Occlusion Handing

Kang Li[1,2,3], Fa-Zhi He[1,2,*], *Senior Member*, *CCF*, and Hai-Ping Yu[1,2]

[1]*State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072, China*
[2]*School of Computer Science, Wuhan University, Wuhan 430072, China*
[3]*School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China*

E-mail: likang@hubu.edu.cn; {fzhe, seaping}@whu.edu.cn

Received November 15, 2016; revised July 12, 2017.

Abstract Visual tracking is an important area in computer vision. How to deal with illumination and occlusion problems is a challenging issue. This paper presents a novel and efficient tracking algorithm to handle such problems. On one hand, a target's initial appearance always has clear contour, which is light-invariant and robust to illumination change. On the other hand, features play an important role in tracking, among which convolutional features have shown favorable performance. Therefore, we adopt convolved contour features to represent the target appearance. Generally speaking, first-order derivative edge gradient operators are efficient in detecting contours by convolving them with images. Especially, the Prewitt operator is more sensitive to horizontal and vertical edges, while the Sobel operator is more sensitive to diagonal edges. Inherently, Prewitt and Sobel are complementary with each other. Technically speaking, this paper designs two groups of Prewitt and Sobel edge detectors to extract a set of complete convolutional features, which include horizontal, vertical and diagonal edges features. In the first frame, contour features are extracted from the target to construct the initial appearance model. After the analysis of experimental image with these contour features, it can be found that the bright parts often provide more useful information to describe target characteristics. Therefore, we propose a method to compare the similarity between candidate sample and our trained model only using bright pixels, which makes our tracker able to deal with partial occlusion problem. After getting the new target, in order to adapt appearance change, we propose a corresponding online strategy to incrementally update our model. Experiments show that convolutional features extracted by well-integrated Prewitt and Sobel edge detectors can be efficient enough to learn robust appearance model. Numerous experimental results on nine challenging sequences show that our proposed approach is very effective and robust in comparison with the state-of-the-art trackers.

Keywords visual tracking, convolutional feature, gradient operator, online learning, particle filter

1 Introduction

Visual tracking has a wide range of applications such as intelligent surveillance, human interaction and virtual reality[1]. Although many tracking algorithms have been studied in recent years, it is still a challenging problem to form a robust tracker because of illumination change[2-3], occlusion[4], deformation and rotation[5]. In order to deal with such problems, researchers focus on exploiting observation model, such as boosting[6-7], structure SVM[8], sparse representation[9-13] and subspace learning[14], which can be divided into two categories: generative model and discriminative model.

In general, generative trackers typically learn an appearance model by a generative process and then search for the most similar target according to reconstruction error. IVT[14] incrementally learns a low dimensional PCA subspace representation, which can effectively model smooth pose variation. L1 tracker[9] assumes that the target could be represented by a sparse linear combination of target templates and trivial templates. After solving an l1-regularized least squares

Regular Paper

This paper is supported by the National Natural Science Foundation of China under Grant No. 61472289 and the National Key Research and Development Project of China under Grant No. 2016YFC0106305.

*Corresponding Author

problem, the candidate sample with the smallest projection error is selected as the target. Through a tracking decomposition scheme, VTD[15] shows that the target can be represented by a linear combination of object templates and trivial templates. Compared with discriminative algorithms, such generative trackers often get more accurate location when the target changes smoothly.

On the other hand, discriminative trackers usually train a classifier to separate the target from the background. For classifier training, these trackers often crop patches near the target location as positive samples and crop patches far away from the target location as negative samples. Finally, such trackers select the sample with the maximum classification score as the target. Thanks to the improvement of machine learning, several sophisticated algorithms have been applied to tracking, such as boosting, SVM and Bayesian. OAB[6] tracker uses multiple instance learning instead of traditional supervised learning and adapts the classifier while tracking the object. Moreover, it selects the most discriminating features for tracking resulting in stable tracking results. STRUCK[8] presents a framework for adaptive tracking based on structured output prediction. It uses a kernelized structured output support vector machine with a budgeting mechanism as appearance model for real-time applications. CT[16] trains appearance model based on the features which are extracted from multi-scale image feature space. It builds an online update naive Bayesian classifier to separate the target from the background, which leads to a real-time and accurate tracker. Compared with generative trackers, discriminative trackers are more robustness because purely generative trackers cannot handle complicated background well.

2 Related Work

To form a robust tracker, Wang *et al.*[1] stated that the features used in tracking systems play the most important role. A good feature can significantly improve the tracking performance even with a simple classifier.

There are many hand-crafted features used in tracking algorithms, such as local binary patterns[17], haar-like features[18-19], contour features[20-23] and other descriptors[24]. Recently, convolutional neural network (CNN)[25-27] is widely used in many image processing fields such as image classification, object recognition and visual tracking. In their convolutional layer, features are extracted by convolving with several filters, which are trained from raw data offline with little human intervention.

However, although these convolved features are robust, the filter training process is time consuming and the model trained using a large amount of general data may not suit tracking specified target. To overcome these weaknesses, we propose to use well integrated first-order derivative edge gradients instead of CNN filters to extract convolutional features.

On one hand, for object tracking, the initial targets always have clear contour features which are robust to illumination change. On the other hand, in image processing, first-order derivative edge gradients are often used to extract contour by convolving them with images. Among these first-order derivative edge gradients, Prewitt and Sobel operators are two complementary gradient operators. Related studies in facial expression detection show that the Prewitt operator is more sensitive to horizontal and vertical edges while the Sobel operator is more sensitive to diagonal edges[28-29].

Fig.1 presents a car under different light conditions (red rectangles) and corresponding contour features (green rectangles). In this figure, the contour features are gained by convolving target image with different Prewitt and Sobel operators. The yellow ovals indicate the similar structure of these contour features, which demonstrates that these features remain stable even under different light conditions. Therefore, in this paper, we integrate several Prewitt and Sobel operators to extract convolved contour features which form the appearance model of the target. The main contributions are listed below.

1) A novel appearance model based on multiple contour features is proposed. To extract contour features, we carefully select and integrate several Prewitt and Sobel operators to convolve with the target image. These contour features are robust to illumination change, which makes our algorithm able to track the target under complex light conditions.

2) We analyse and show that the bright parts of the contour features could provide more important information than dark parts. Therefore, we propose to measure the similarity between candidate samples and to only use bright pixels in our trained appearance model. Because occlusions are always smooth and have few contours, this strategy could exclude obstructed areas which help to solve partial occlusion problem.

3) After evaluating the target location at the end of each frame, a corresponding incrementally update method is presented to adapt appearance change, which shows robustness in our experiment.