Abstract:
Visual tracking is one of the fundamental problems in computer vision. Its numerous applications include robotics, autonomous driving, augmented reality and 3D reconstruction. In essence, visual tracking can be described as the problem of estimating the trajectory of a target in a sequence of images. The target can be any image region or object of interest. While humans excel at this task, requiring little effort to perform accurate and robust visual tracking, it has proven difficult to automate. It has therefore remained one of the most active research topics in computer vision.

In its most general form, no prior knowledge about the object of interest or environment is given, except for the initial target location. This general form of tracking is known as generic visual tracking. The unconstrained nature of this problem makes it particularly difficult, yet applicable to a wider range of scenarios. As no prior knowledge is given, the tracker must learn an appearance model of the target on-the-fly. Cast as a machine learning problem, it imposes several major challenges which are addressed in this thesis.

The main purpose of this thesis is the study and advancement of the, so called, Discriminative Correlation Filter (DCF) framework, as it has shown to be particularly suitable for the tracking application. By utilizing properties of the Fourier transform, a correlation filter is discriminatively learned by efficiently minimizing a least-squares objective. The resulting filter is then applied to a new image in order to estimate the target location.

This thesis contributes to the advancement of the DCF methodology in several aspects. The main contribution regards the learning of the appearance model: First, the problem of updating the appearance model with new training samples is covered. Efficient update rules and numerical solvers are investigated for this task. Second, the periodic assumption induced by the circular convolution in DCF is countered by proposing a spatial regularization component. Third, an adaptive model of the training set is proposed to alleviate the impact of corrupted or mislabeled training samples. Fourth, a continuous-space formulation of the DCF is introduced, enabling the fusion of multiresolution features and sub-pixel accurate predictions. Finally, the problems of computational complexity and overfitting are addressed by investigating dimensionality reduction techniques.

As a second contribution, different feature representations for tracking are investigated. A particular focus is put on the analysis of color features, which had been largely overlooked in prior tracking research. This thesis also studies the use of deep features in DCF-based tracking. While many vision problems have greatly benefited from the advent of deep learning, it has proven difficult to harvest the power of such representations for tracking. In this thesis it is shown that both shallow and deep layers contribute positively. Furthermore, the problem of fusing their complementary properties is investigated.

The final major contribution of this thesis regards the prediction of the target scale. In many applications, it is essential to track the scale, or size, of the target since it is strongly related to the relative distance. A thorough analysis of how to integrate scale estimation into the DCF framework is performed. A one-dimensional scale filter is proposed, enabling efficient and accurate scale estimation.