Reliable Object Pose Estimation

In this thesis, we address reliable estimation of an object's position and orientation, called the object's pose, from one or more images. In case of symmetries, occlusions, etc., potentially many poses may explain the observed image, and a single pose estimate cannot explain this ambiguity. Instead, we we aim to estimate the ambiguity as a pose distribution to enable reliability for downstream robot systems.

First, we consider pose estimation for objects in an industrial case, where the relevant pose uncertainties can be described by a distribution over a small discrete set of rotations, which are easily represented by a deep learning model. Uncertainty quantification is especially interesting in this case, since erroneous pose estimates could lead to damaged products or equipment, and we show that the probabilistic models allows us to reliably avoid failure.

Second, a common approach to pose estimation is establishing correspondences between points in the image and points on the object. Usually, it is assumed that an image point could only correspond to one point on the object, however, that assumption breaks in case of ambiguities such as those imposed by symmetries. We present SurfEmb, modeling correspondence distributions, and showing how to use the distributions to obtain better pose estimates. Our method was on top of the main pose estimation benchmark for almost a year.

Third, single-view pose estimation inherently suffers from depth ambiguity and sensitivity to occlusions. To that end, we present EpiSurfEmb which finds a pose which explains the correspondence distributions across views. Additionally, we also combine the image-object correspondence distributions from SurfEmb with epipolar geometry to estimate scene-object correspondence distributions. Our results show that, using multiple images, we can reduce errors by 80-91 %.

Fourth, we present Ki-Pode, which flips the correspondence distribution problem to estimate distributions of the projection of predefined keypoints on the object surface, and we show how these distributions can provide an estimate of the pose distribution. Due to the lack of a way to normalize over the pose space, we only show distribution estimation on the rotation space, where we provide more reliable estimates across objects than other methods.

Fifth, we present SpyroPose which addresses how to scale distribution estimation models to the six-dimensional pose space. The main idea is to learn distributions at multiple resolutions, allowing more efficient training and many orders of magnitude fewer evaluations at test time due to a sparse evaluation of the pose space. The method can be applied to both rotation and pose space. We present state-of-the-art results on rotation distributions estimation, and on pose distribution estimation, we present the first qualitative results on real images and the first quantitative results at all. We also show that the method extends readily to a multi-view version, presenting a principled way to fuse pose information from multiple images.