

論文の内容の要旨

論文題目 Part-level Shape, Pose and Kinematics

Understanding of Man-made Articulated Objects

(関節物体に対するパーツレベルの形状・姿勢・動作に関する理解)

氏 名 川名 雄樹

This thesis proposes an end-to-end, unified framework for understanding articulated objects from a single-view RGBD input. The primary focus lies in the estimation of part shape, pose, and kinematic parameters for man-made articulated objects in a scene, such as household appliances and furniture. These objects are assumed to have a base and one or more connected movable parts, excluding loop structures, separations, or structures where one movable part is attached to another.

The proposed method addresses multiple challenges associated with the understanding of articulated objects captured from a single viewpoint. The task is highly ill-posed, in terms of shape reconstruction and kinematics estimation due to partial observation of shapes and the static target without motion. The diversity and combination of part poses, shapes, kinematics, part counts, and structures presents a significant challenge in understanding 3D attributes of the articulated objects due to the arising complexity. Moreover, the need for 3D part-level supervision signals, such as kinematic parameters and part labels, makes training low data-efficient. Existing studies have failed to provide a comprehensive solution, by assuming prior knowledge for the target articulated object to manipulate for obtaining prior kinematics information, limiting part structure and counts for the target articulated objects to limit the complexity in shape reconstruction, and requiring a whole set of 3D part-level annotations for all training data.

This thesis proposes a comprehensive pipeline in response to the above

problems. The pipeline takes a single RGBD image, camera intrinsics, and, optionally, foreground masks as input, and outputs the part shape, pose, kinematic parameters of individual parts, and the hierarchical structure of parts that make up the instance. This thesis further delves into unsupervised learning for data efficiency besides the supervised approach, by exploiting the fact that certain everyday articulated objects, such as ovens and washing machines, tend to have consistent part structures. The method also explores unsupervised segmentation of parts into finer semantic shapes, such as the handle of the doors, which is essential to recognize a preferable contact point for motion planning. By combining both the supervised and unsupervised approaches, the proposed pipeline is learned in a semi-supervised manner, which is the most realistic setting that we have access to the annotation for a portion of data while reducing the amount of required annotation for better data efficiency whenever possible.

In summary, this thesis makes significant strides in understanding articulated objects from a single-view input, developing a unified framework that handles part detection, shape reconstruction, pose estimation, kinematic estimation, and segmentation of the finer shape details. Through this work, we hope to facilitate further exploration and advancement in the understanding of man-made articulated objects.