Pose Estimation of Kinematic Chain Instances via Object Coordinate Regression

Frank Michel
Frank.Michel@tu-dresden.de
Alexander Krull
Alexander.Krull@tu-dresden.de
Eric Brachmann
Eric.Brachmann@tu-dresden.de
Michael Ying Yang
Ying.Yang1@tu-dresden.de
Stefan Gumhold
Stefan.Gumhold@tu-dresden.de
Carsten Rother
Carsten.Rother@tu-dresden.de

TU Dresden
Dresden
Germany

Accurate pose estimation of object instances is a key aspect in many applications, including augmented reality or robotics. For example, a task of a domestic robot could be to fetch an item from an open drawer. The poses of both, the drawer and the item have to be known by the robot in order to fulfil the task. 6D pose estimation of rigid objects has been addressed with great success in recent years. In large part, this has been due to the advent of consumer-level RGB-D cameras, which provide rich, robust input data. However, the practical use of state-of-the-art pose estimation approaches is limited by the assumption that objects are rigid. In cluttered, domestic environments this assumption does often not hold. Examples are doors, many types of furniture, certain electronic devices and toys. A robot might encounter these items in any state of articulation.

This work considers the task of one-shot pose estimation of articulated object instances from an RGB-D image. In particular, we address objects with the topology of a kinematic chain of any length, i.e. objects are composed of a chain of parts interconnected by joints. We restrict joints to either revolute joints with 1 DOF (degrees of freedom) rotational movement or prismatic joints with 1 DOF translational movement. This topology covers a wide range of common objects (see our dataset for examples). However, our approach can easily be expanded to any topology, and to joints with higher degrees of freedom.

Figure 1: **Our dataset.** These images show results on our dataset. The estimated poses are depicted as the blue bounding volume, the ground truth is shown as the green bounding volume of the object parts.

Articulated Pose Estimation. To estimate the pose of a kinematic chain $\hat{H} = (H_1, \ldots, H_K)$ we need to find the 6D pose H_k for each part k. The problem is however constrained by the joints within the kinematic chain. Therefore, we can find the solution by estimating one of the transformations H_k together with all 1D articulations $\theta_1 \ldots, \theta_{K-1}$, where θ_k is the articulation parameter between part k and $k+1$. We assume the type of each joint and its location within the chain to be known. Given θ_k we can derive the rigid body transformation $A_k(\theta_k)$ between the part k and $k+1$. The transformation $A_k(\theta_k)$ determines the pose of part $k+1$ as follows: $H_{k+1} = H_k A_k(\theta_k)^{-1}$. We can use this to estimate the 6D poses of all parts and thus the entire pose $\hat{H}$ of the chain from a single part pose together with the articulation parameters.

Hypothesis Generation. We use a random forest to produce two kinds of predictions for each pixel i. Given the input depth image, each tree in the forest predicts object probabilities and object coordinates for each separate object part of the kinematic chains (Fig. 2, middle). An articulated pose hypothesis is sampled as follows. We draw a single pixel i_1 from the inner part ($k = 2$) randomly using a weight proportional to the object probabilities $p_k(i)$. We pick an object coordinate prediction $\mathbf{y}_k(i_1)$ from a randomly selected tree t. Together with the camera coordinate $\mathbf{x}(i_1)$ at the pixel this yields a 3D - 3D correspondence $(\mathbf{x}(i_1), \mathbf{y}_k(i_1))$. Two more

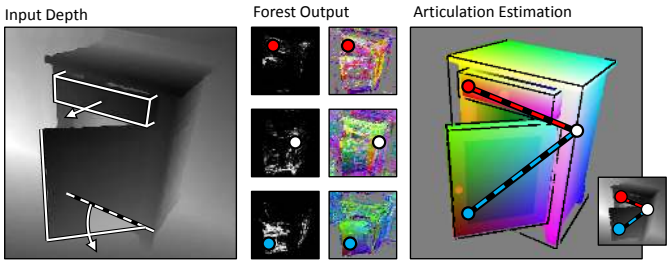

Figure 2: **Articulation estimation.** Left: Input depth image, here shown for the cabinet. The drawer is connected by a prismatic joint and the door is connected by a revolute joint (white lines are for illustration purposes). Middle: Random forest output. Top to bottom: Drawer, base, door, where the left column shows part probabilities and the right the object coordinate predictions, respectively. Right: Articulation estimation between the parts of the kinematic chain using 3D-3D correspondences between the drawer / base and door / base. Note that the three correspondences (red, white, blue) are sufficient to estimate the full 8D pose.

correspondences $(\mathbf{x}(i_2), \mathbf{y}_{k+1}(i_2))$ and $(\mathbf{x}(i_3), \mathbf{y}_{k-1}(i_3))$ are sampled in a square window around i_1 from the neighbouring kinematic chain parts $k+1$ and $k-1$. We can now use these correspondences to estimate the two articulation parameters θ_{k-1} and θ_k between part k and its neighbours. We derive $A_k(\theta_k)$ and $A_{k+1}(\theta_{k+1})$ and map the two sampled points $\mathbf{y}_{k+1}(i_2)$ and $\mathbf{y}_{k-1}(i_3)$ to the local coordinate system of part k. We have now three correspondences between the camera system and the local coordinate system of part k, allowing us to calculate the 6D pose H_k using the Kabsch algorithm. The 6D pose H_k together with the articulation parameters yields the pose $\hat{H}$ of the chain. Fig. 2 illustrates this process

Results. We contribute a new dataset consisting of over 7000 frames annotated with articulated poses of different objects: two cupboards, a laptop and a toy train[1]. The objects show different grades of articulation ranging from 1 joint to 3 joints. When compared to the 6D pose estimation pipeline of Brachmann *et al.* [1] our method shows superior results (89% averaged over all sequences and objects) in comparison to the baseline (29%). Qualitative results are shown in Fig. 1. Employing articulation constraints within the kinematic chain results in better performance on the individual parts as well as for the kinematic chains in its entirety. Our approach of pose sampling for kinematic chains does not only need less correspondences, it is also robust when dealing with heavy self occlusion.

Conclusion. We present a method for pose estimation of kinematic chain instances from RGB-D images. We employ the constraints introduced by the joints of the kinematic chain to generate pose hypotheses using K 3D-3D correspondences for kinematic chains consisting of K parts.

[1] E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton, and C. Rother. Learning 6d object pose estimation using 3d object coordinates. In *ECCV*, 2014.

[1]This dataset will be part of the ICCV 2015 pose challenge: http://cvlab-dresden.de/iccv2015-pose-challenge