Evento	Salão UFRGS 2019: SIC - XXXI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2019
Local	Campus do Vale - UFRGS
Título	A CNN Method for Background Subtraction and 3D Pose Estimation Based on Depth Images
Autor	MATHEUS ALAN BERGMANN
Orientador	CLAUDIO ROSITO JUNG

Title: A CNN Method for Background Subtraction and 3D Pose Estimation Based on Depth Images

Author: Matheus Alan Bergmann **Advisor:** Claudio Rosito Jung
Institute: Federal University of Rio Grande do Sul - UFRGS

Introduction

Pose estimation is one of the most important tasks in computer vision. It aims to relate two different coordinates systems, where typically one of them is the camera coordinate system and the other the world coordinate system (which can also be associated with another camera).

In this work, we present an end-to-end Convolutional Neural Network (CNN) model that both subtracts the background and regresses the pose of the object. It assumes the input image contains the object roughly in the center of the frame and with the camera placed in a position that makes most of the image be occupied by the item whose pose we want to estimate. Also, we assume the object will stay with his top side pointing to the top of the image or with a small rotation angle. The output is a single normalized vector representing the rotation (2-DoF).

Data Representation

In this work, we propose to represent the pose as a unit vector that points from the object center to the camera position, which reduces the error in the angles close to 0º and 359º. It presents only two degrees of freedom (2-DoF) since camera roll is not considered. However, it can be extended to 3-DoF by predicting the *up vector* of the virtual camera, which is orthogonal to the direction vector and encodes roll information.

Loss Function

We use a hybrid loss function that allows us to start training the segmentation part first and then refine it while we train the pose estimation part. This loss function is parametrized by an α weight that ponderates segmentation and pose estimation. A larger value for α means the optimizer should prioritize reducing the segmentation loss over the pose loss, while the opposite is valid for smaller values of α.

Experiments

1) One key issue on the proposed model is the α weight that controls the hybrid loss function pondering segmentation and pose estimation. We compare 3 different α values: $\alpha = 0.2$, $\alpha = 0.8$ and $\alpha = 0.5$ which represent the method prioritizing pose, segmentation and treating both equally, respectively.

2) To measure the effectiveness of the segmentation net over the final result, we trained the network only with the alignment part. The architecture for pose estimation was the same, but instead of the input being a segmented depth image, this model was trained and tested using depth images containing both the object and the background.

3) Since we have synthetic meshes, we can train our model from scratch by rendering a huge amount of images. However, the segmentation network is trained first, and we performed an experiment to evaluate if the initial layers of the segmentation network could be used as an initialization for the pose estimation network.

Results

Our angular error at the 90th percentile is around 10º, which means that 90% of our errors are below 10º.