6 DOF POSE ESTIMATION OF A CAR FOR AUTONOMOUS DRIVING SYSTEM

Suma S[1], Lavanya P[1], Sampreetha S[1], Raksha R[1], Mrs. Nikitha[2]

[1]Student, Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.
[2]Associate Professor, Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.

---***---

Abstract – A conceptual framework for 6DoF object pose estimation for autonomous driving scenario is demonstrated. The demonstrated method perfectly detects traffic participants in a RGB image while simultaneously regressing their 3D translation and rotation vectors. The method, called 6D-VNet, extends Mask R-CNN by adding customized heads for predicting vehicle's rotation and translation. Convolutional Neural Network is also used to estimate the 3D translation of an object. The paper displays the use of translational relapsing in the joint losses is important for the 6DoF pose estimation task, where object translation distance along longitudinal axis varies significantly. In addition, there is inclusion of the mutual information between traffic participants via a modified non-local block.

***Key Words*: 6DoF, CNN, R-CNN**

1. INTRODUCTION

If we need to achieve autonomous operation of a vehicle in urban situations with unpredictable traffic, their flaws must be overcome by accurately estimating pose in 3 dimensional space. Object pose inference aims to detect objects and estimate their orientations and translations. For the same reason computer vision can be made to gain high-level perceptive from digital images. 6 degrees of freedom refers to the specific number of access that a rigid body is able to freely move in 3-D space (rotation and translation in 3D). CNN method interprets to Convolutional Neural Networks which is a very popular algorithm for image classification and typically consists of convolution layers, activation function layers, pooling (primarily max-pooling) layers to decrease dimensionality without losing a lot of features

2. LITERATURE SURVEY

The problem in Autonomous driving system is similar to industrial robots in which robots should be able of recognizing the position and orientation of a part before grasping it. A monocular-based 6-degree of freedom (DOF) pose analysing process which allows robots to seize large-size parts at informal poses is simulated. A camera was mounted on the robot end-flange and oriented to measure several featured points on the part before the robot moved to seize it. experimental tests, the part poses estimated with the method described in this paper were compared with those produced by a laser tracker. Robotic intelligent grasping tests were also successfully performed in the experiments.[1].

Evaluating the 6D-pose is provocating due to the variety of objects as well as the complexity of a scene caused by clutter and occlusions between objects.. The 3D rotation of the object was evaluated by relapsing to a quaternion representation. They also introduced a novel loss function that allow Pose CNN to handle symmetric objects. Pose CNN is highly vigorous to occlusions and provide accurate pose estimation using only color images as input. [2].

We suggest a single-shot approach for concurrently furnishing an instance in an RGB image and predicting its 6D pose without requiring multiple stages or having to examine multiple hypotheses. The important constituent of this approach was CNN architecture that directly predicted the 2D image locations of the projected vertices of the object's 3D bounding box. The object's 6D pose was then estimated using a PnP algorithm. For post-processing, a pose refinement step can be used to boost the accuracy.[3].

The provocation of 6DoF pose evaluation from a single RGB image under vigorous occlusion or truncation is surpassed by using a Pixel-wise Voting Network (PVNet) to regress pixel-wise unit vectors marking to the keypoints and use these arrays to vote for key point locations using RANSAC. This creates a flexible representation for localizing occluded or truncated key points.. Experiments states that the proposed method surpassed the state of the art on the LINEMOD, Occlusion LINEMOD and YCB Video datasets by a large margin, while being efficient for real-time pose estimation.[4].

In this approach determination of traffic participants in a monocular RGB image is done concurrently relapsing their 3D translation and rotation vectors. The method, called 6D-VNet, broadens Mask R-CNN by adding customised heads for predicting vehicle's finer class, rotation and translation. It was shown that the inclusion of translational regression in the joint losses is crucial for the 6DoF pose estimation task, where object translation distance along longitudinal axis varied significantly.[5].

3. CONCEPTS

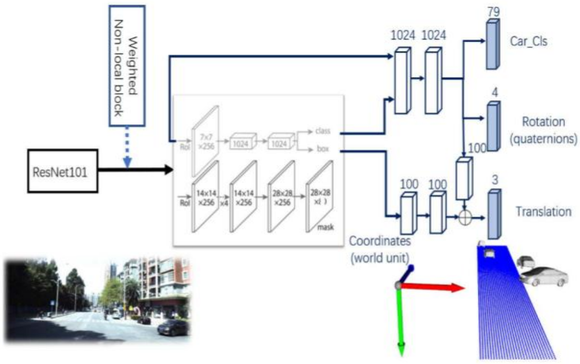

Fig -1: System pipeline: 6D-VNet takes a monocular image as input and performs the vehicles' 6DoF estimation. The grey box serves a canonical instance division network and the dark blue branch is for estimating object 6DoF pose and its sub-category.

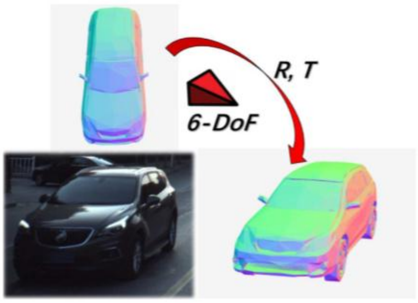

Fig -1: The important constituents of the proposed network: following the detection head in 2D space, the 6D-VNet regresses the 6DoF rotation and translation vector simultaneously in 3D space.

4. TECHNOLOGIES

SYSTEM CALLIBRATION METHOD ,IMAGE PROCESSING METHOD, This method is designed to be applied on the production line with large part intelligent grasping requirements and adjust the robot's path to adapt to the pose of the deviated part.[1]

POSE CNN, a convolutional neural network for 6D object pose evaluation. PoseCNN disjoins the evaluation of 3D rotation and 3D translation. Two new loss functions are introduced for rotation evaluation, with the ShapeMatch -Loss designed for symmetric objects. As a result, PoseCNN is handling occlusion and symmetric objects in cluttered scenes. We also introduce a large scale video dataset for 6D object pose estimation.[2].

Approaches such as Viewpoints and Keypoints and Render for CNN, cast object categorization and 3D pose estimation into classification tasks, specifically by seperating the pose space. In contrast, PoseNet suggests using a CNN to directly relapse from an RGB image to a 6D pose[3].

VOTING BASED KEYPOINT LOCALIZATION, PVNet predicts pixelwise object labels and unit vectors that represent the direction from every pixel to every keypoint. We generate hypotheses of 2D locations for that keypoint as well as the confidence scores through RANSAC-based voting. [4].

The approach, called 6D-VNet, broadens Mask R-CNN by adding customised heads for predicting vehicle's finer class, rotation and translation. The output will precisely estimates the vehicle's voxels in 3D occupancy.[5].

5. CONCLUSION

This problem is quite challenging from many perspectives: Object detection under severe occlusions, Truncation of objects, Variations in lighting and appearance, Cluttered background objects, Dataset has significant image noise, Camera pose estimation (self-localization problem).

To overcome these problems, many technologies and new updated ideas are used. The steps that is followed are:

3-D Object Detection, Localization, Object classification used to produce class labels for the object, Instance Segmentation, 6-Dof pose estimaton (transational and rotational vectors).

6. REFERENCES

[1] "Tao Liu, Yin Guo *, Shourui Yang, Shibin Yin and Jigui Zhu"."Monocular-Based 6-Degree of Freedom Pose Estimation Technology for Robotic IntelligentGrasping Systems", 2017.

[2] "Yu Xiang1;2, Tanner Schmidt2, Venkatraman Narayanan3 and Dieter Fox"."PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes", 2018.

[3] "BugraTeki,Sudipta N.Sinha,PascalFua"." Real-Time Seamless Single Shot 6D Object Pose Prediction",2018.

[4] "SidaPeng, Yuan Liu, Qixing Huang, HujunBao, Xiaowei Zhou"."PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation",2018.

[5] "Di Wu, Zhaoyong Zhuang, Canqun Xiang, Wenbin Zou and Xia Li"." 6D-VNet: End-to-end 6DoF Vehicle Pose Estimation from Monocular RGB Images",2019.