A CAMERA-BASED TRACKING SYSTEM FOR DISTRIBUTION NETWORK INSPECTION BASED ON UNMANNED AERIAL VEHICLES

Zhai Ruicong	Chen Hao	Zhang Feng	Xu Zhihai	Yang Chengcheng
China	China	China	China	China
zhairuicong@163.com	chengzi_zone@163.com	89264020@qq.com	writetohai@163.com	yccio@qq.com

ABSTRACT

A Camera-based Tracking System for Distribution Network Inspection Based on Unmanned Aerial Vehicles The emerging technology of unmanned aerial vehicle (UAV) has become more affordable and practicable for distribution network inspections. In this paper, we present and demonstrate a camera-based tracking algorithm that achieves accurate in-air localization for distribution network inspection， In order to Reduce accuracy of GPS positioning requirements. Specifically, our approach consists of Three stages. It first generate the route using GPS information of towers. Following that, the GPS information of towers automatic update by the camera-based SLAM algorithms. According to the new GPS information of towers, the yaw maneuvering command was generated and route dynamic planning on the basis of the pre-determined route. Meanwhile, the ideal camera shooting angle for Distribution Network lines and towers is calculated, Then the control angle can be adjusted automatically to realize the tracking process. While a number of localization methods are possible (most notably avehicle differential GPS (DGPS) system), we chose to work with a camera-based system for the ubiquity of cheap, low weight and high quality camera systems means that such systems have ideal characteristics for UAVs,And the feasibility of this algorithm is proved by tests.

1. INTRODUCTION

In order for unmanned aerial vehicle (UAV) to operate safely and effectively alongside humans in distribution network inspection, they must be aware of their surroundings. One aspect of this awareness is knowledge of the Tower 3D position and orientation of objects in the scene,This knowledge is important to perform find-and-photograph of towers. In this work we focus on towers for which a prior training time to learn the appearance and shape of the towers is allowed. Our goal is to infer the 3D pose of such towers, in clutter, from a single RGB image in real time for the purpose of enabling the UAV to inspect towers of distribution network.

While deep neural networks have been successfully applied to the problem of object detection in 2D [1, 2, 3], they have only recently begun to be applied to 3D object detection and pose estimation[4, 5, 6,7]. Unlike 2D object detection, it is prohibitive to manually label data for 3D detection. We propose to overcome such limitations by using synthetically generated data. Synthetic data has been gaining traction in recent years as an efficient means of both training and evaluating DNNs for computer vision problems for which collecting ground truth data is laborious.

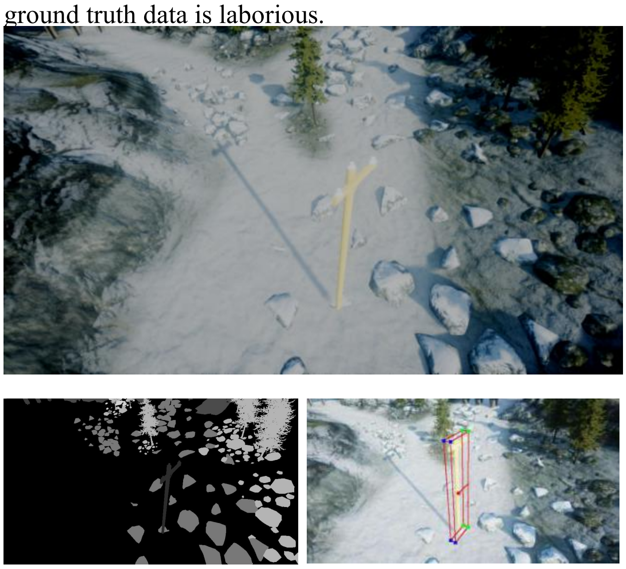

Figure 1. The Tower dataset was generated by placing 3D household object models in virtual environments. Each snapshot consists of a stereo pair of RGB images (only one of which is shown, top),pixelwise segmentation of the objects (bottom left), depth (not shown), 2D/3D bounding box coordinates (bottom right), and 3Dposes of all objects (not shown).

We believe that 3D detection and pose estimation fall within this category and are thus a natural fit for synthetic data due to the difficulty of acquiring accurate ground truth.

In this paper we use the towers dataset, consisting of a large number of snapshots for training and evaluating robotics scene understanding algorithms in outdoor environments. Specifically, as shown in Fig. 1, each snapshot consists of a stereo pair of image frames with corresponding depth images, along with the 3D poses, 2D/3D bounding boxes, projected 3D bounding boxes, and pixelwise segmentation of the known objects in the scene. Our contributions are thus as follows:

• Using one-shot, deep neural network-based system that infers, in near real time, the 3D poses of towers in outdoor scenes from a single RGB image without requiring IMU.

• An integrated UAV system that shows the estimated poses are of sufficient accuracy to solve real-world tasks such as inspect power towers and power line path following.

2. APPROACH

We propose a two-step solution to the problem of detecting and estimating the 6-dof pose of all instances of a given set of power towers from a single RGB image. First, the depth neural network estimates the two-dimensional confidence mapping of key points of all objects in the image coordinate system. Secondly, the peaks of these belief graphs are fed back to the standard perspective n-point (PnP) algorithm [8] to estimate the 6-dof pose of each object instance. In this section, we describe these steps and our new approach to generating synthetic data for training neural networks.

2.1Data generation

How to generate effective training data for network is a key problem in this study. Compared to 2D object detection, where marking borders is relatively easy to annotate, 3D object detection requires marking data that is almost impossible to generate manually.
While data can be semiautomatically tagged (using tools such as LabelFusion[9]), the labor-intensive nature of the process hinders the ability to generate training data with sufficient variation. For example, we do not know the posture estimation of any real-world training data for 6-dof objects, including extreme lighting conditions or posture.
To overcome these limitations of real data, we turn to synthesized data. Specifically, we use a combination of non-photorealistic random field (DR) data and photorealistic data to take advantage of both. These two types of data are complementary and produce much better results than if they were used alone, as shown in the subsequent experiments. Synthetic data has the additional advantage of avoiding overfitting the distribution of a particular data set, resulting in a network that is robust to changes in light, camera, and background.
All data is generated by a custom plug-in called NDDS[10] Unreal Engine 4(UE4). Using asynchronous, multithreaded sequential frame fetching, the plug-in generates data at a rate of 50-100 Hz, which is much faster than the default UE4 screenshot or publicly available Sim4CV tool [11]. Figure 1 shows an example of an image generated using UE4 plug-in, illustrating the diversity of field random data and realistic data.

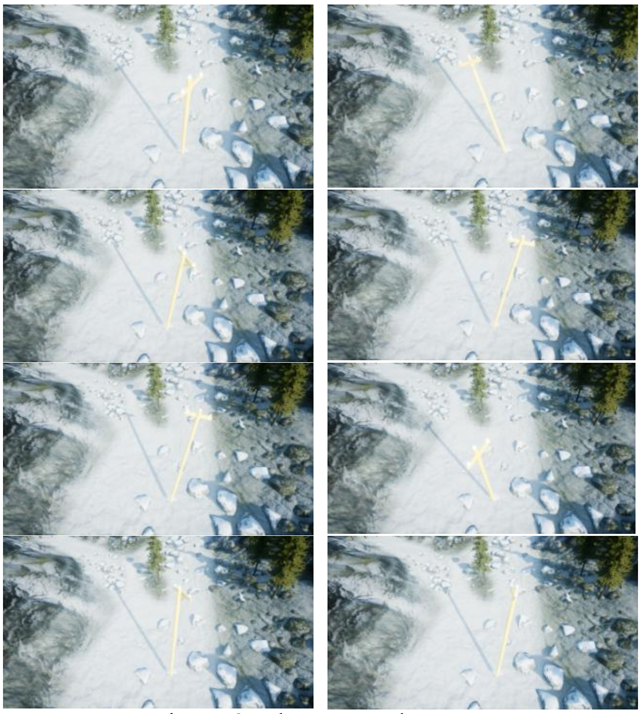

Figure 2. The Tower dataset.

2.2 TRAINING

In the training, we used 60 k domain-randomized image frames and ~ 60 k realistic image frames. While we use a single group of ~ 60 k images in front (one data set for each object). For PoseCNN[5], we use publicly available weights.
The system was trained on an NVIDIA DGX workstation (consisting of four NVIDIA P100 or V100 gpus) and tested using NVIDIA Titan X.

3.3 UAV MANIPULATION

For our purposes, the ultimate test of a pose estimation method is whether its accuracy is sufficient for UAV control.

Figure 3. The UAV.

We attached a intel RealSense D435 camera to the top of UAV. Use Pixhawk2 and NVIDIA Jetson TX2 companion computer to contorl uav for distribution network inspection.
The Jetson TX2 mounted on the J120 Carrier board. The TX2 has a GPU for machine learning and USB3 with sufficient throughput for depth cameras.
For quantitative results, we find 4 towers of distribution network inspection in outdoor, The UAV was instructed to move to a shoot a close-up point

above the tower, then execute a top-down take a photo, yielding 10 trials per tower. Of these 10 attempts, the number of successful grasps were as follows: 10 (first tower), 10(second tower), 10 (third tower), 9 (Fourth). These experiments demonstrate that our DOPE pose estimation method is robust enough in real-world conditions to execute successful move to a shoot a close-up point and take a photo.

3. CONCLUSION

We have presented a Camera-based Tracking System for Distribution Network Inspection Based on Unmanned Aerial Vehicles.this system can reduce accuracy of GPS positioning requirements, and operate uav safely and effectively alongside humans in distribution network inspection.

REFERENCES

[1] J. Dai, Y. Li, K. He, and J. Sun. R-FCN: Object detection via region-based fully convolutional networks. In NIPS, 2016.

[2] J. Redmon and A. Farhadi. YOLO9000: Better, faster, stronger. In CVPR, 2017.

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In ECCV, 2016.

[4] M. Rad and V. Lepetit. BB8: A scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth. In ICCV, 2017.

[5] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. In RSS, 2018.

[6] B. Tekin, S. N. Sinha, and P. Fua. Real-time seamless single shot 6D object pose prediction.In CVPR, 2018.

[7] T. To, J. Tremblay, D. McKay, Y. Yamaguchi, K.Leung, A. Balanon, J. Cheng, and S. Birchfield.NDDS: NVIDIA deep learning dataset synthesizer.https://github.com/NVIDIA/Dataset_Synthesizer, 2018. 2

[8] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An accurate O(n) solution to the PnP problem. International Journal Computer Vision, 81(2), 2009.

[9] P. Marion, P. R. Florence, L. Manuelli, and R. Tedrake. LabelFusion: A pipeline for generating ground truth labels for real RGBD data of cluttered scenes. In ICRA, 2018.

[10] T. To, J. Tremblay, D. McKay, Y. Yamaguchi, K. Leung, A. Balanon, J. Cheng, and S. Birch-field. NDDS: NVIDIA deep learning dataset synthesizer, 2018. https://github.com/NVIDIA/Dataset_Synthesizer.

[11] M. Mueller, V. Casser, J. Lahoud, N. Smith, and B. Ghanem. Sim4CV: A photo-realistic simulator for computer vision applications. International Journal of Computer Vision, pages 1–18, 2018.