1-14

Classifying Overhead Images Taken by Different Height Cameras Predicted Depth by Fully Convolutional Residual Networks with Convolutional Neural Network

Shu SUMIMOTO　Yuichi MIYATA　Ryuta YOSHIMURA
Yoko UWATE　Yoshifumi NISHIO
(Tokushima University)

1. Introduction

Drones are recently infiltrating various fields, for example delivery, rescue, guard and so on. It is necessary that drones fly safely. Then image recognition by deep learning is becoming important for that drones fly safely. However, standing people in overhead images from the view of drones are not able to be recognized by using YOLOv3 which is an object detection algorithm.

In this study, we investigate the prediction of the depth of some objects, such a human and chairs, in overhead images with Fully Convolutional Residual Networks (FCRN). We classify the AI-generated predicted depth images and we aim at differentiating human or other objects.

2. Proposed System

We propose to classify predicted depth images with FCRN. First, we predict depth of overhead RGB images of a human, a chair and car taken by monocular cameras in Figs. 1 and 2.

Second, we classify the RGB images and the predicted depth images of a human and a chair or a car with a convolutional neural network which has 2 convolutional layers, 2 pooling layers and 2 fully connected layers.

We compare the learning and test accuracies when a camera is close to an object with when a camera is far from an object. When they are close to a camera, we classify images of a human and a chair. On the other hand, when they are far from a camera, we classify images of a human and a car.

Figure 1: Overhead RGB image.

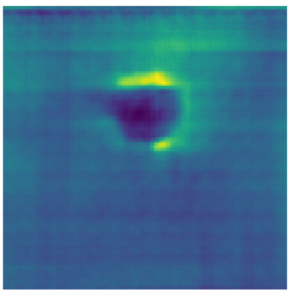

Figure 2: Predicted depth image.

3. Simulation results

We define as the learning steps 300, the number of the training data sets 60 which are 40 overhead images of a human and 20 overhead images of a chair and the number of the test data sets 10 which are 5 overhead images of a human and 5 overhead images of a chair. They are taken by a camera which is close to an object. Table 1 shows average of learning and test accuracies when we classify only RGB images 10 times and we classify only predicted depth images 10 times.

We define as the learning steps 100, the number of the training data sets 80 which are 40 overhead images of a human and 40 overhead images of a car and the number of the test data sets 10 which are 5 overhead images of a human and 5 overhead images of a car. They are taken by a camera which is far from an object. Table 2 shows average of learning and test accuracies when we classify only RGB images 10 times and we classify only predicted depth images 10 times.

From Tables 1 and 2, when a camera is close to an object, the test accuracy of classifying RGB images is higher than predicted depth images. On the other hand, when a camera is far from an object, the test accuracy of classifying predicted depth images is higher than RGB images.

Table 1: Average of learning and test accuracies when a camera is close to an object.

	RGB images	Predicted depth images
Learning accuracy	0.9650	0.9016
Test accuracy	0.87	0.72

Table 2: Average of learning and test accuracies when a camera is far from an object.

	RGB images	Predicted depth images
Learning accuracy	0.8335	0.7649
Test accuracy	0.38	0.58

4. Conclusion

From these simulation results, we consider it is effective for classifying images taken by a camera which is far from an object to predict the depth of images.

However, the learning accuracies of classifying predicted depth images are lower than RGB images. Then we will try to raise accuracies of classifying predicted depth images.