Recognition of Door Handles to Enable a Robot to Open Doors

1. Introduction

In the last few decades the desire for robots to perform human like characteristics has led to the development of numerous algorithms for robot vision, learning and navigation. Many recent studies are focused on generating a map of the environment and self exploration problem. These works enable to obtain the pose of a robot as well as a map of its surroundings. Yet, they do not consider the problem of accessing new locations in an indoor structure by manipulating doors. Since mobile robots have an unlimited work space, their capabilities should be improved to let them access to new areas without any human assistance. For example when a robot is sent into a facility, which may be potentially hazardous for human health, it should be able to navigate throughout the building by itself [1]. To increase the circulation space, recognition of door handles is a key problem to be solved during robot navigation.

In this study the aim is to recognize and localize door handles by both using gray scale 2D images and 3D point cloud images. Visible light laser with a camera which is placed on a STanford Artificial Robot (STAIR) is used as the data source. The laser and a camera operate in collaboration to provide 3D coordinates for each pixel of a 2D image. SVM (Support Vector Machines) classification method is applied to both 2D and 3D images so that 2 different classifiers are generated to identify the door handles. The localization of a handle in a new given image is performed by a sliding window search. We initially evaluate the decision boundary for the 2D-classifier on a regular lattice and obtain the possible regions of the handle on the image. Then we prune these results by evaluating the 3D-classifier on those regions. In other words the initial estimates that are obtained by 2D images are improved by using the 3D data. The results show that for opaque doors the localization accuracy is high, but false positives prevent accurate localization for transparent doors.

2. Data Collection and Feature Extraction

Visible light laser with a camera, which is mounted on a STAIR, is used for collecting training and test data (Figure 1). The camera and visible light laser are coupled together so that for each pixel of an image, *X-Y-Z* coordinates can be acquired. We have collected data from about 40 different doors and formed a data set that contains a total of 70 gray scale images and their corresponding point clouds. Both opaque and transparent door images are taken in order to evaluate the performance of our method realistically.

Extraction of the positive and negative data samples is performed by the help of 2D images since they are easier to work with instead of the 3D point clouds. The process is semi automatized by asking the user to locate only the center of the region of interest on a 640 x 480 image. For example, while generating a positive training data, once the center of the door handle is selected by the user, a fixed sized window (101 x 61) is positioned around that point. Pixel values and the corresponding X-Y-Z coordinates for each pixel in the window are then lined up to generate our feature vectors for the 2D and 3D classifiers respectively.

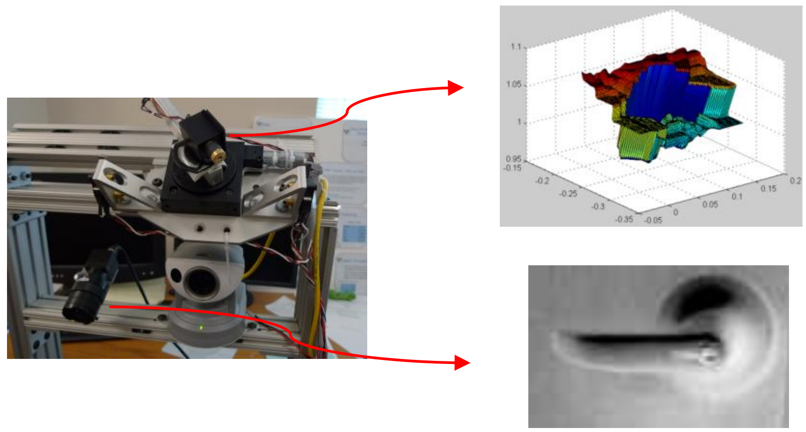

Figure 1 – Visible light laser with a camera is used to obtain the image of a door handle and the corresponding 3D point cloud view. Image pixel values and corresponding X-Y-Z coordinates are used as feature vectors.

3. Algorithm

We first obtain the decision regions for both 2D and 3D classifiers. SVM with a linear kernel is used for training both classifiers. To determine the location of a door handle on a given image we need to apply a sliding window search on the entire image. Direct application of this method on a point cloud data would take 3 times more time than a gray scale image. So, initially the 2D classifier is employed and probable regions of the handle are obtained. Then those regions are classified again by the 3D classifier. So that false positives of the 2D classifier are pruned. Only if the 2D classifier cannot predict the position of the handle at a particular case (i.e. the handle may not be recognized when the image is dark), we apply a complete sliding window search on the point cloud image.

4. Performance of Support Vector Machines

Prior to SVM logistic regression and least squares classification methods are employed, but their error rate was higher than SVM. When we apply SVM to a training set containing 50 positive and 500 negative samples, the classification error for both 2D and 3D classifiers are given in Table 1 and Table 2 respectively. Test set contains 20 positive and 200 negative samples. Instead of using equal number of negative and positive training samples their ratio is selected as 10 since an input door image normally contains only a single handle which covers a much smaller area then the negative data regions.

Table 1 – SVM training and test errors for gray scale image data.

2D	False Positives %	False Negatives %
Training	0.0	0.0
Test	1.2	35.0

Table 2 – SVM training and test errors for 3D point cloud data.

3D	False Positives %	False Negatives %
Training	0.0	0.0
Test	2.0	0.0

Table 1 gives us a general idea about the classification performance. The high rate of false negatives is caused by the dark images that are used for testing. In these cases 2D image is not informative enough to recognize the handle, but instead point cloud data can be used. In fact, Table 2 shows that all the test samples are successfully recognized since false negative percentage for the 3D classifier is 0.0%. Still complicated point clouds of transparent doors cause the 2% false positive rate to be unavoidable.

5. Performance of Sliding Window Search

We have extracted overlapping detection windows on a regular lattice (sliding windows) and evaluated the decision boundaries given by SVM. The difference between the consecutive search windows is set to 10 pixels. Taking into account the morphology of the robot and the height at which the camera is mounted we have a priori information about the position of the robot arm in the images. This knowledge is used to eliminate the false positives caused by the robot arm.

The results for an opaque and transparent door are shown in Figure 2. Although the handle is detected in both cases, false positives could not be eliminated completely when the door is transparent. Actually having higher error rate with a glass door which yields a much more complicated 3D point cloud image compared to an opaque door that yields a mostly flat 3D shape was expected. Moreover the number of training samples corresponding to the transparent regions might not be sufficient enough to enable an accurate classification.

We define the localization error of a handle as the pixel-wise difference between the ground truth (i.e. user selected) center of the handle (on 2D image) and the center predicted by the sliding window search. The average localization error is obtained as 4.84 pixels for the opaque doors in the test set. For the transparent doors the localization error is obtained as 67.12 pixels. Hence we can conclude that for opaque doors the handle can be accurately localized; however we cannot obtain the true position of the handle if the door is transparent. Although the sequential evaluation of the 2D-3D classifiers reduces the number of false positives significantly, accurate localization necessitates a complete elimination of the false positives. For better accuracy a separate classifier for the transparent doors might be trained.

(a) **(b)**

Figure 2 – Sliding Window results for (a) opaque and (b) glass doors

However in that case we would also need a pre-classifier which differentiates the types of the doors. Clustering the final predicted centers by k-means classifier (with k = 2) according to their depth values may also help to eliminate false positives.

6. Conclusion and Further Work

In this study we have shown an implementation of SVM for the recognition of door handles. We could reliably obtain the center of a door handle for opaque doors by using the prior information about the position of the robot arm and sequential usage of 2D, 3D classifiers. Although using both images and point clouds helps to reduce the number of false positives, we cannot guarantee to eliminate them for transparent doors. Hence localization of the handle for those cases could not be achieved successfully.

A superior approach may be training separate classifiers for opaque and transparent doors, but that would also necessitate training of a pre-classifier that differentiates the types of doors. Another way to reduce the number of false positives might be adding k-means clustering as a final step.

Once the localization of the handle is performed, the axis of rotation and the orientation (left or right turn) of the handle can also be determined. Furthermore force and torque that should be applied by the robot can be reduced by determining the position of the tip of the handle. This will not only increase power efficiency, but also enhance success rate of opening the door.

7. Acknowledgement

Special thanks to Ellen Klingbeil for inspiration and fruitful discussions.

8. References

1. Ellen Klingbeil, Ashutosh Saxena, Andrew Y. Ng, Learning to Open New Doors, (2008)