Classifying Objects From a 3D Model

Kayla Bonnstetter
Computer Science
St. Olaf College
Northfield, Minnesota 55057
bonnstet@stolaf.edu

Adrian Rossing
Computer Science
St. Olaf College
Northfield, Minnesota 55057
rossing@stolaf.edu

Abstract

This paper's focus is on classifying individual faces of objects that make up a three dimensional model, given their coordinates in three-dimensional space and their color properties. There is plentiful research for the classification of objects from 2D images or 3D point clouds. Our method borrows from these techniques. It uses the projected 3D coordinates of an object from 2D stereoscopic images to find geometric features and the object's RGB color values within that area. The randomForest classifier in R creates three initial predictions and histogram comparison tool validates those predictions.

We observed 90.96% accuracy for the first prediction and 98.88% accuracy for all three. When testing the 93 identified doors in our dataset (4 types), 94.62% appeared in the first prediction and 98.92% appeared in the top three. As evident from our results, this method of classification works well to categorize objects from a 3D model.

Introduction

While extensive progress has been made to automate computer vision in the 2D scope and in 3D with the use of point clouds, other techniques for object recognition appear to require additional research. In the field of 2D object identification, door recognition has provided numerous techniques for detecting doors in an image. This has been done both by using color boundaries in images and by observing physical characteristics with lasers. Prominent experiments use canny edge detection and CORNER [1] and the AdaBoost algorithm [2] in detection of doors. Research completed in 3D space often incorporates the use of 3D point clouds to identify objects. As in door recognition, certain features from an object of interest are used for classification. These include features such as an object's shape or where it is in relation to other objects [3]. We used an approach based on 2D images to find the features of interest within the model being created. This paper explains the techniques used to classify object faces that have been projected from 2D stereoscopic images.

Methods

Data used in conjunction with our research was produced by hand using a software package called Eriol, developed at St. Olaf College. Object faces were created from stereo images of hallways from St. Olaf's science center, Regents Hall, using superpixels. The faces, as they appear in the images, are called contours. Contours were joined with matching contours in their image pair (and other images the contour appeared in) and placed in 3D space, giving them a set of 3D coordinates. These pairs of contours are called tiles. Once we receive information on tiles, we use its given 3D coordinates and the objects' color properties to classify objects. In our data sets we use 38 different types of objects that are likely to appear, such as different types of doors and walls, as well as smaller objects such as fire alarms.

The first four features we use for classification are height, width, area, and the average RGB values of an object. Because objects usually sit perpendicular to the floor, we only use the y values of an object to determine height. We subtract the lowest y value from the highest y value to do this. While objects tend to be oriented to the y-axis, this is not the case with the x-axis, especially when looking down halls. Thus, we use the Euclidean distance formula on the highest and lowest x and z values.

We then take a simple approach to finding area; we use the product of the width and height as though the object were a rectangle and do not consider any holes it may have or indents to its outline. Average RGB values were calculated from all pixels residing within the tile boundaries.

These features are used to produce three initial classifications using the RandomForest package in R. They are assigned ranked values of .5 for the most likely prediction, .333 for the second, and .167 for the final prediction. These numbers were chosen because they add up to 1 and assigned weights to the predictions. The learning dataset we developed has 783 distinctive objects, consisting of 38 different object types. This dataset retrains the classification package each time to increase accuracy, but is not necessary when optimizing for efficiency. With three potential object classifications, the names are further rated using a modification of the OpenCV program Histogram Comparison.

The histogram comparison stage allows us to more accurately match the colors present in a tile to an object's "known" color properties. Instead of comparing histogram values, the program we developed creates a histogram for the tile trying to be classified. It then compares it with histograms from images identified as each of the three initial classifications. The modified OpenCV Histogram Comparison program computes a correlation (a floating point number between 0 and 1) based on the similarities between the histogram for the known object and the histogram for the object being identified. The comparison images come from a data set containing 206 images that was selected by hand to accurately represent a given object. Finally these correlations are multiplied by a coefficient of .5 and added to the ranked values for the predictions to calculate the order of the final prediction along with a numerical value.

Results

Tests were performed on our dataset containing the known identity of objects. There were 1,968 tiles included which consisted of 47 different objects. We randomly split the known tiles in half. The first half was used as our training set (984 tiles and 47 objects) while the second half was tested (984 tiles and 39 objects). RandomForest included 47 different learned objects and the dataset of images used to create histograms included images for all objects. Our method correctly classified 90.96% (895 tiles) in the first prediction from our program. Between the three given predictions, 98.88% (973 tiles) included the correct object name. We observed 1.12% (11 tiles) whose actual name did not appear in any of the three predictions. Further tests included adjusting the coefficient. The results suffered with both larger and smaller coefficients. For instance, changing the constant to 1 resulted in naming 83.77% correctly in the first prediction and having the correct name appear in the three predictions 94.24%.

The recognition of tiles that are doors is very successful with our methods. There are four different types of doors in the tiles included. With a total of 93 doors tested, 94.62% (88 tiles) were correctly classified by the first prediction and 98.92% (92 tiles) of the correct names appeared in one of the three predictions. Only 1 door was not included in the top three predictions.

Discussion

The tests we ran delivered consistent results. These results fall within the standard we observed in our research, allowing us to conclude that our methods of classification should be used to name objects in situations where contours in 2D images are combined to create tiles with color information and coordinates in 3D space. (For example, other research has been completed in object recognition using histograms with detection rates between 70% and 90% [4]. Established 2D object recognition techniques result in between 72% [2] and 98% accuracy [1].) We completed research to understand previous techniques in classifying doors. Although many of the established methods occurred in 2D space, our results from stereoscopic pairs of images fall within the normal ranges for similar research. We are unfortunately unable to compare our results to other 3D door recognition results due to the differences in approach, but again believe that our results provide accurate, usable results.

We also tested to see if changing the formulas we use to determine height and width would affect the measurements for a given object. The second height formula we use is similar to equation 1. Instead of using the highest and lowest x values though, we used y values. We also tried a formula similar to our original height formula in which we subtracted the lowest x value from the highest x value. We thought that this would change the height and width values, but it did not.

To obtain accurate results, solid testing data must be procured. There were limitations that, if avoided, may have allowed for further research in the time allotted. We were restrained due to a lack of an existing dataset. Development of the data used in our research in the classification of objects in Regents Hall was completed while attempting to build a functional model. With an existing model in place, object recognition would have been simpler due to a larger data set of known objects.

Future Work

We envision our program being used to assist in the classification of objects in Eriol and other real instances. When implemented within these programs, users will be given the three predictions in their final order. Clicking on the correct classification will result in a number of things happening. First and foremost, the tile being named will adopt the correct name. (Incase the correct name is not provided, users may type in the correct name). Second, the learning data set used in R will be appended with a new line containing the correctly identified object and its dimensions and average RGB values. Finally, a portion of the image will be added to the image in the data set associated with the object classified. All of these considerations will allow for a classification technique that incorporates the ability to adjust for a changing data set and expand when new items are encountered.

Because our goal was to find an initial way to recognize objects with the type of data we had, there are many areas for future research. Since our classification method relies on correct dimensions and colors, two areas that will be important to continue studying are how to deal with occlusion and imperfect colors. Currently tiles are flagged by hand when they are troublesome. Automation of this process should help improve accuracy. Another way to improve results would be to include other identifying features, such as which plane an object is in. For example, doors and walls are in vertical planes while floors and ceilings are in horizontal planes. Additionally, we were not able to look at how an object's relation to similar, or particular objects, could help recognition. For example, a door is normally contained within a wall. Using a more robust approach to determining area may also be helpful, especially for smaller objects and objects with holes.

Conclusion

We have discussed our methods for classifying object faces in a 3D model derived from 2D stereoscopic images. We introduce two classification techniques that provide consistent and accurate results on the data tested. A learned classifier in R, randomForest, provides initial predictions while comparing histograms finalizes the predictions. The experiment results show that our solution can match and in some cases outperform existing object classification techniques.

References

[1] Tian, Y., Yang, X., Arditi, A. *Computer Vision Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons*. 2010. Web.

[2] Hensler, J., Blaich, M., Bittel, O. *Real Time Door Detection Based on AdaBoost Learning Algorithm*. 2010. Web.

[3] Golovinsky, A., Kim, V.G., Funkhouse, T. *Shape-based Recognition of 3D Point Clouds in Urban Environments*. 2009. Web.

[4] Murillo, A., Košecká, J., Guerrero, J., Sagüés , C. *Visual Door Detection Integrating Appearance and Shape Cues*. 2007. Web.