Remote Surface Classification for Robotic Platforms

Will Roderick, Connor Anderson, Aaron Manheim

wrtr@stanford.edu, connora@stanford.edu, manheima@stanford,edu

Motivation

As robots move from controlled lab environments into the outside world, situational awareness and adaptability in uncertain and varying conditions are becoming increasingly important. In many cases, robots must be able to detect, interact with, and navigate around physical surfaces. One such application is air to surface transitions for aerial robots, which can prolong mission life for imaging, inspection, and physical data collection [1],[2].

(Above) "Perching UAV" from Stanford BDML

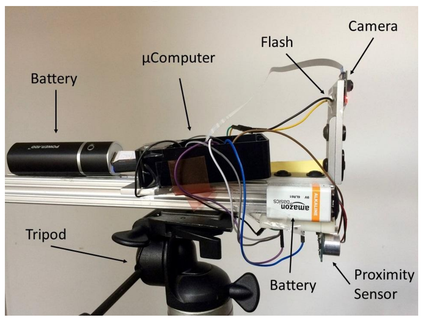
A wide range of machine learning techniques have been applied to texture classification, including bag of visual words and convolutional neural network models [3],[4]. These projects have typically used solely raw RGB photos as inputs. Small robots are typically equipped with other sensors in addition to a camera for navigation in uncontrolled environments, which we hypothesized we could use to improve image classification results.

In this project, we augmented vision with additional sensing modalities, including light reflection and distance to the surface, to enable a learning algorithm to classify surface textures in varying conditions with computational cost suitable for a microprocessor. We focused on glass, indoor walls, exterior walls, tree bark, and foliage.

Materials & Methodology

Experimental Setup

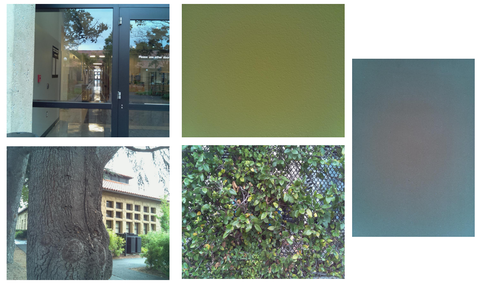
To collect data, we constructed a device with an integrated camera, proximity sensor, and flash which were controlled by a Raspberry Pi microcomputer. When prompted, the system automatically took a picture both with and without flash in under a second. At the same time, the distance from the surface was recorded using the proximity sensor and the data was stored on the microcomputer along with a timestamp.



Data Collection

Image Categorization

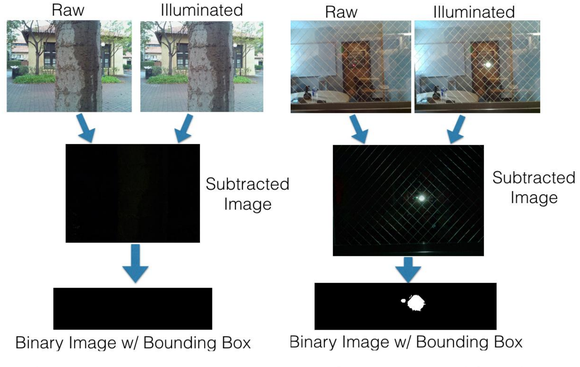
Using the aforementioned setup, we collected over 1000 images of 5 surfaces: foliage, tree bark, glass, building exterior, and building interior (representative figures below). These images are sorted for classification with both the flash and no-flash versions. Note the similarity between indoor and outdoor walls.



Above: Sample Images of Textures (Glass, Indoor Wall, Outdoor Wall, Tree Bark, Foliage)

Features

To improve the performance of our learning algorithm, we extract unique features from the images, which relate only to the subtraction of a raw RGB image from a flash enhanced image (i.e. which cannot be obtained from the raw image alone). By subtracting the raw image from the illuminated image, we then can apply a normalization, conversion to greyscale, and thresholding to convert the image into a binary image (see Figures below - note that not all images look this clean). From here we experiment with the following parameters: maximum region area, number of distinct regions, ratio of largest area to second largest area, pixel dispersion from largest area, brightest pixel intensity, and distance to surface.



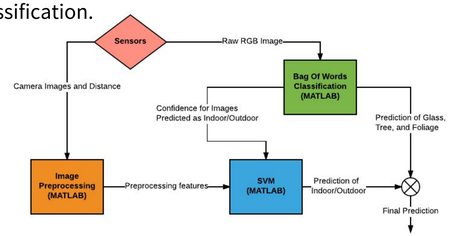
Above: Preprocessing Sequence for a Tree and for Glass.

By training our learning algorithm on these features and removing features one at a time to see the effect on the generalization error, we can determine which features contribute most to the success of the algorithm.

Models

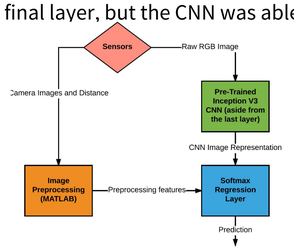
1. Modified Bag of Words (Modified BoVW)

Raw images were fed to a BoVW classification algorithm implemented in Matlab, which then fed the images of the categories that were hardest to classify to an SVM for more accurate classification.



2. Convolutional Neural Network (CNN)

Raw images were used to train the last layer of the Inception V3 neural network. We had planned to feed in the image processing features in the last layer with the CNN image representation before the final layer, but the CNN was able to perform with high accuracy even without the sensor augmentation.



The algorithm iterates to reduce cross-entropy, the equation for which is given here:

$$H_{y'}(y) = -\sum_i y'_i log(y_i)$$

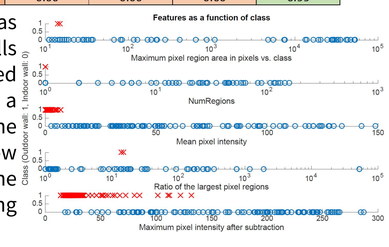
Results – Modified BoVW

The BoVW was able to classify foliage, bark and glass with 99% or more accuracy. However, indoor walls were mislabeled as outdoor walls 10% of the time.

Confusion Matrix on Raw Images

	Foliage	Indoor Walls	Outdoor Walls	Bark	Glass
Foliage	1.00	0.00	0.00	0.00	0.00
Indoor Walls	0.00	0.88	0.10	0.02	0.00
Outdoor Walls	0.00	0.03	0.96	0.00	0.01
Bark	0.00	0.00	0.00	0.99	0.01
Glass	0.01	0.00	0.00	0.00	0.99

Images classified as indoor or outdoor walls were then preprocessed and fed through a Kernelized SVM. The plots to the right show the results of the features created during image processing.



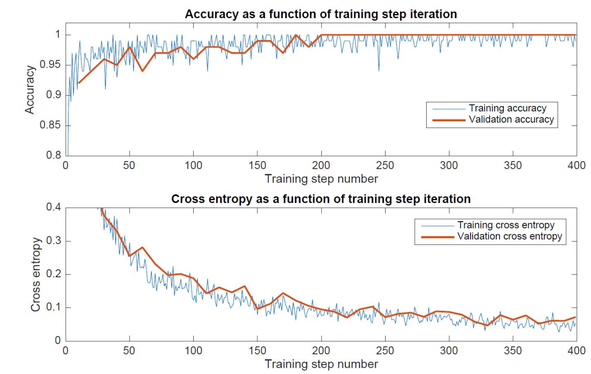
The SVM was able to classify indoor walls with 96% accuracy (an 8% improvement shown below).

Modified BoVW Confusion Matrix

	Foliage	Indoor Walls	Outdoor Walls	Bark	Glass
Foliage	1.00	0.00	0.00	0.00	0.00
Indoor Walls	0.00	0.96	0.02	0.02	0.00
Outdoor Walls	0.00	0.00	1.00	0.00	0.01
Bark	0.00	0.00	0.00	0.99	0.01
Glass	0.01	0.00	0.00	0.00	0.99

Results - CNN

Given the raw RGB frames for the 5 classes of interest, the CNN was able to achieve 99.9% accuracy.



Discussion & Future Work

Discussion

Overall, image enhancement with additional sensor modalities improved the accuracy of the BoVW model. Specifically, reflectance properties improved classification performance on indoor and outdoor walls. On the other hand, the neural network was able to achieve ~99.9% accuracy with just the raw image frames. These results indicate that while the simple model may not be as accurate as a complex neural network, the simple model, which runs 6 times faster, is suitable for robotic surface classification. These results support that optimizing this model or a simpler neural network could enable real-time classification on small robotic processors.

Future work

Over the next weeks and months, we plan to use this system for robotic applications, including enabling aerial robots to detect viable landing locations. Specifically, we plan to optimize the simpler model for real time classification on a small processor. Farther into the future, we will focus on adaptability in cluttered environments as well as including a wider range of surfaces and ambient conditions into the classification scheme.

Acknowledgements

We would like to thank Professors Andrew Ng and John Duchi as well as our TA Francois Germain.

References

[1] Roderick, William R. T., Mark R. Cutkosky, and David Lentink. "From touchdown to takeoff: at the interface of flight and surface locomotion." *Interface Focus*. 2016. In Press.

[2] Desbiens, Alexis Lussier, Alan T. Asbeck, and Mark R. Cutkosky. "Landing, perching and taking off from vertical surfaces." *The International Journal of Robotics Research*. (2011): 0278364910393286.

[3] Andrearczyk, Vincent, and Paul F. Whelan. "Using Filter Banks in Convolutional Neural Networks for Texture Classification." *arXiv* preprint arXiv:1601.02919 (2016).

[4] Zhang, Jianguo, Marcin Marszałek, Svetlana Lazebnik, and Cordelia Schmid. "Local features and kernels for classification of texture and object categories: A comprehensive study." *International journal of computer vision*. 73, no. 2 (2007): 213-238.