Autonomous Computer-Vision-Based Human-Following Robot

Akshay Gupta, Nick Gloria

{akshaygu, ngloria}@stanford.edu

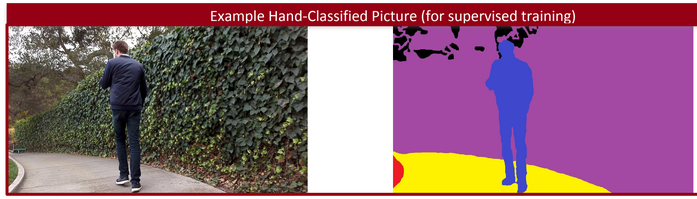
Introduction

- For “golf caddy” product: autonomous golf bag carrier
- Computer vision directs robot to follow human, navigate terrain
- Enhances golfer experience, assists elderly or disabled golfers
- Robot has single camera, onboard processor
- Terrain classification problem: classifying image into traversable/untraversable terrain (avoiding green, sand traps, etc.)
- Target-following problem: identify and track target, maintaining distance

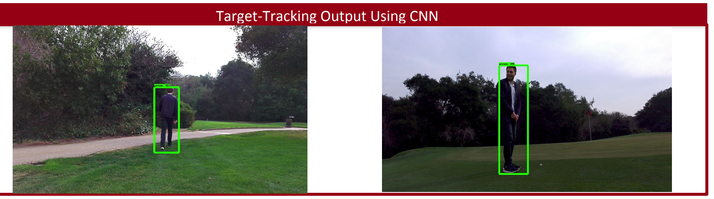
Data and Features

- Dataset collected via onboard camera (Jetson SoC)
- Training and test data sets selected from images taken during product testing
- Supervised data provided by hand-classified, color-coded images

Example Hand-Classified Picture (for supervised training)

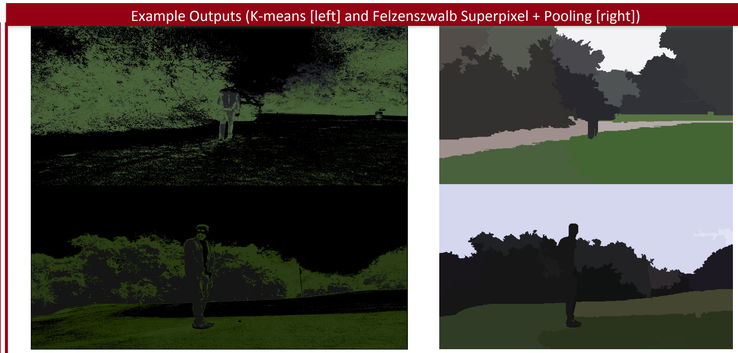


Target-Tracking Output Using CNN



Example Outputs (Convolutional Softmax Regression)

Example Outputs (K-means [left] and Felzenszwalb Superpixel + Pooling [right])



Models

Both unsupervised and supervised methods were implemented to this classification problem.

Target Tracking: Convolutional Neural Network - MobileNet SSD

- Using pre-trained neural network, identified following target with bounding box

Image Classification: K-means-based Segment Detection

- K means clustering on a test images resulting in color based clusters
- Clusters picked for appropriate terrain areas, which generates segments based on norm distance in the color space

Image Classification: Felzenszwalb Superpixel Segmentation + Pooling

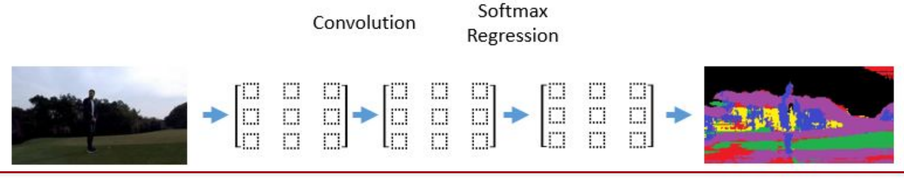
- SciKit based image segmenter, results in boundary generation based on pixel vise difference both locally, and globally after clustering
- Draws boundaries after generating clusters segmenting into large regions
- The pixel values are averaged in each cluster and repainted onto image

Image Classification: Convolutional Layer

- Square pixel stencil converts each pixel to a vector of pixels
- Linear combination of neighboring pixel vectors gives access to gradient

Image Classification: Softmax Regression

- Uses key-painted images for training
- Classifies each pixel into one category
- Classes: green, traversable grass, bunker, target, path, etc.



Results and Discussion

The results show that with given data, the algorithms implemented can roughly classify the robot images, with some shortcomings and pitfalls.

Convolution + Softmax Regression

- Rough accuracy with noise, limited by single-layer classification
- Suffers from low number of hand-classified images for training

Image Classification: K-means-based Segment Detection

- Pixelated images generated with area segmentation and limited closed curves
- Cluster selection needs tuning and automation to increase efficiency

Image Classification: Felzenszwalb Superpixel Segmentation + Pooling

- Good clustering but fails to captures all boundaries, also generates more clusters than required

Target Tracking: Convolutional Neural Network - MobileNet SSD

- High accuracy and precise detection, unable to detect smaller or thin objects
- Needs to be fine tuned on golf course setting, including equipment and obstacles

Future Steps

- Generation of larger dataset on golf course images for training the networks
- Use edge-detection algorithm to implement pooling layer (i.e. use most common classification to classify identified segment)
- Extension of single classification to multiple classification layers (neural network)
- Devise efficient means of classifying images (e.g. augmentation by edge detector, stamping) to allow for expansion of training set

References

[1] Felzenszwalb, P.F. & Huttenlocher, D.P. International Journal of Computer Vision (2004) 59: 167. https://doi.org/10.1023/B:VISI.0000022288.19776.77

[2] Howard, et al. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.” [Astro-Ph/0005112] A Determination of the Hubble Constant from Cepheid Distances and a Model of the Local Peculiar Velocity Field, American Physical Society, 17 Apr. 2017, arxiv.org/abs/1704.04861.