Campus Location Recognition using Audio and Visual Signals

James Sun, (Teammate TBD)
SUNetID:Jsun2015
Email: jsun2015@stanford.edu

I. Introduction

This project aims to create a system that can recognize (classify) geographical locations on the Stanford University Campus using Audio and Video signal inputs. The system will leverage labeled training data to train a classifier. This project will try to avoid focusing on sophisticated Image Processing techniques such as RANSAC/SIFT/SURF feature recognition, and instead utilize large quantities of simple features. Also, this project will attempt to emphasize audio based recognition as most previous work in this topic has been in computer vision.

II. Related Work

A previous CS229 course project identified landmarks based on visual features [1]. [2] gives a classifier that can distinguish between multiple types of audio such as speech and nature. [3] investigates the use of audio features to perform robotic scene recognition. [4] takes a empirical approach toward recognizing environmental sound.

III. Data

Data is expected to be collected using a smartphone that has built-in audio and visual recording capabilities and has GPS capabilities. Training data will be labeled either by GPS data or by hand. Visual data will be captured by camera in still frames. Capturing video data by use of a GoPro^{TM}camera is also possible.

IV. Methods

The goal is to have the system recognize natural geographic aggregates rather than recognize individual fine-grain coordinates. For example, the system will label a data set as belonging to "The Quad" or "Bytes Cafe", similar to how a person would naturally describe an environment. I expect to use supervised learning algorithms to separate data points based on audio and visual features.

A. Audio Features

As audio is potentially the more interesting data type, I have come up with a few basic features to evaluate. These include the following:

- Frequency Spectrum Bandwidth
- Frequency Spectrum Variance
- Frequency Spectrum mean
- Intensity variance
- Mean Intensity

References

[1] A. Crudge, W. Thomas, and t. . Kaiyuan Zhu.
[2] L. Chen, S. Gunduz, and M. T. Ozsu, "Mixed type audio classification with support vector machine," in *2006 IEEE International Conference on Multimedia and Expo*, July 2006, pp. 781–784.
[3] S. Chu, S. Narayanan, C. c. J. Kuo, and M. J. Mataric, "Where am i? scene recognition for mobile robots using audio features," in *2006 IEEE International Conference on Multimedia and Expo*, July 2006, pp. 885–888.
[4] S. Chu, S. Narayanan, and C. C. J. Kuo, "Environmental sound recognition with time and frequency audio features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1142–1158, Aug 2009.