

Plankton Image Classification

Rui Wang (ruiw1)

Problem Statement

- How to classify grayscale images of Plankton into one class out of 121?

Overall Experiment Design

- Because the whole dataset is fairly large, I choose a subset in reasonable size that can work on my laptop. (10,000 images in 28 classes).
- Choose SVM model because of properties of typical image features.
- 80% images used as training set, left ones are test set.

Experiment #1

- Choose all the pixels and create a wide matrix
- Didn't work because feature vector is so large. Laptop cannot finish computing within reasonable period of time.

Experiment #2

- Still use pixels as feature, but try to reduce the dimension of feature vector.
- How? For every 50 pixels, get the minimum, maximum and average float values. Reduce the feature vector to $N / 50 * 3$, where N is the total number of pixels in image.

Accuracy and Error Analysis

- **40.001%** accuracy on 2142 test images.
Incredible High 😊
- But pixel feature is not good. It cannot express the essence of image, so it is possible introduce noise.

Experiment #3

- Use more powerful feature – SIFT
- Basic step
 - Extract SIFT key points from all images
 - K-means to cluster them
 - Create histogram of SIFT points distribution for each image. Use this histogram as feature vector.

Accuracy

2060 test images categorized into 28 classes

Number of cluster	Accuracy
100	54.03%
200	56.99%
500	54.99%

Have 16% improvement!

Experiment #4

- How about Histogram of Oriented Gradient feature (Detect Object)?

Accuracy

2131 test images categorized into 28 classes

Length of HOG feature	Accuracy
54	31.44%
216	19.34%
864	18.11%

Experiment #5

- How about combine HOG and SIFT feature?

Accuracy

2060 test images categorized into 28 classes

# of SIFT feature	Length of HOG feature	Accuracy
20	54	30.58%
40	54	30.24%
100	54	28.54%

Error Analysis

- Looks like combining doesn't work. Unknown reason for now, but maybe bagging could help.

What's Next

- Bagging Strategy. Let models trained by different features vote.
- Different models. Like neural network.