

# Handwritten Character Classification

Shengxin Qian  
ECE 590 Text Analysis

November 7, 2016

## 1 Program Work flow

1. Load MNIST dataset with mnist helper functions and transform the original 28 \* 28 images into 1 \* 784 vectors.
2. Bootstrap 500 samples from the entire train dataset (60000 samples) for training and bootstrap 500 samples from test dataset (10000 samples) for testing
3. Use Matlab embedded multi-class SVM functions("templateSVM", "fitcecoc", "predict") to implement linear multiclass SVM.

### 4.1 Baseline Approach

Input raw grayscale value of each pixel into SVM. Each sample has 784 features

### 4.2 Feature Extraction

Use "extractHOGFeatures" function to extract HOG features and input preprocessed data into SVM. Each sample has 144 features(default block size and cell size).

5. Repeat 3, 4 steps for 30 times and calculate average error rate.

## 2 Result

As one can see in the Figure 1, the HOG descriptor could extract gradient features. Those features contains the information of edges which is better than treating each pixel independently.

Training Sample Size	Testing Sample Size	Base Line Error	HOG Error
500	500	0.1469	0.686
500	2000	0.1521	0.0665
2000	500	0.0963	0.0397

Table 1: Handwritten Character Classification Result

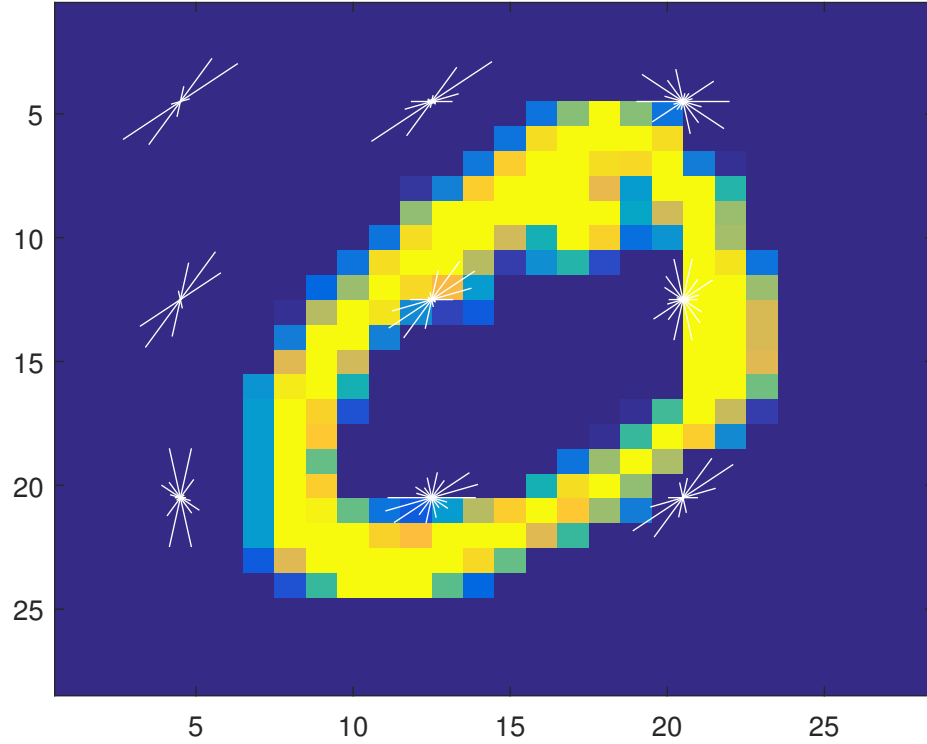


Figure 1: HOG feature over original image