

智能信息处理PJ2

局部特征

具体处理流程

1. 找到特征向量，或者说特征值
 1. 多个局部特征 [opencv tutorial](#)
 1. SIFT
 2. SURF
 3. ORB
2. BoW 方法进行聚类
 1. 默认用 KMEANS 聚类采用欧式距离，得到50个聚类中心
 2. 每张图片用 50 个聚类中心的频率分布表征，一张图片最终 == (1,50)
3. 训练分类器，使用 scikit-learn 中的实现
 1. KNN
 2. SVM
 3. 决策树
 4. 朴素贝叶斯
 5. Logistic 回归
4. 测试
 1. 对不同特征比较
 2. 对不同分类器进行比较

因为 opencv 中 SURF 和 SIFT 算法有版权，要同时用的话需要控制packages的版本

```
pip install opencv-contrib-python==3.4.2.17
pip install opencv-python==3.4.2.17
```

results

特征	SURF	SIFT	ORB
训练集向量总长	(3971087, 64)	(5515132, 128)	(2235245, 32)
SVM	49.87%	53.20%	37.07%
Logistic	40.40%	43.27%	35.60%
Native bayes	30.27%	33.47%	29.67%
Decision tree	26.20%	36.27%	25.13%
KNN	35.60%	40.13%	32.00%

Table1: 模型在测试集上的精度，SVM 最高