1 Summary

1.1 Software Used

Scikit-learn [3]: It is an open source machine learning library for the Python programming language.

NumPy and SciPy [2, 4]: These are numeric and scientific libraries in Python that provide utility methods.

1.2 Feature Extraction

SIFT : Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. We used the SIFT implementation written in C by David Lowe [5]. For each image, this generated a set of feature points each represented by a 128 size vector.

1.3 Similarity/Distance Measures

We used K-Means clustering in which the distance between two vectors is defined as the Euclidean distance between them.

1.4 Classifier

After experimenting with multiple classifiers like naive bayes, random forests, etc., we chose to use a support vector classifier due to its accuracy.

2 Algorithm

We have used the **Bag of Image Words** approach [1] in classifying the images. The steps involved in our algorithm are as follows:

1. For each image in the training data, extract the SIFT feature points (as 128-vectors) using David Lowe's C program for SIFT. These feature vectors are assembled together to form a NumPy matrix of dimensions N x 128 (where N is the total number of features extracted from all training images).
2. To this matrix, we applied K-Means clustering setting the number of clusters to be the square root of the number of features (N). After this stage, we obtained a set of $\sqrt{N}$ clusters each with a corresponding centroid.
3. Each training image's SIFT feature vectors were quantized to the set of centroids obtained above (each feature vector was replaced with its nearest euclidean distance centroid). Thus, each image could now be represented as a frequency vector of the $\sqrt{N}$ centroids as $\{f_{i,j}\}_{j=1}^{\sqrt{N}}$ where $f_{i,j}$ is the frequency of the j^{th} centroid in the i^{th} image.
4. This set of frequency vectors, one for each image (indicating the frequency of each centroid in the image) was used as input to a support vector classifier (SVC), along with the labels data from labels.txt. The SVC classifier was thus trained.
5. Whenever a new image was to be classified, it's SIFT feature vectors were first extracted, quantized and the image was reduced to the centroids frequency vector form as described above. This was then given to the SVC for testing and the predicted class obtained.

On testing this algorithm on the validation set, we obtained an accuracy of 75.6%.

References

[1] Bag of image words. [Online]. Available: http://en.wikipedia.org/wiki/Bag-of-words_model_in_computer_vision

[2] Numpy. [Online]. Available: http://www.numpy.org/

[3] Scikit-learn. [Online]. Available: http://scikit-learn.org/stable/

[4] Scipy. [Online]. Available: http://www.scipy.org/

[5] Sift implementation. [Online]. Available: http://www.cs.ubc.ca/~lowe/keypoints/