I have tested the new program on my desktop (Win7, 64bit, Python 2.7) and Ubuntu (12.04, 64bit, Python 2.7.6). Please download the new program from Assignment folded called Bag-of-Visual-Words-Image. Since most of you use Windows, I use Win7 as an example in the document.
The approach consists of two major steps called learning and generating, represented in the files learn.py and generate.py.
The script learn.py will generate a visual vocabulary using a user provided set of training images. After the learning phase generate.py will use the generated vocabulary to generate visual words vector for any image given to the script by the user.
I have trained the data already. You can download 4 files (including codebook) from the assignment folder to generate a visual words histogram for a test image directly (Please refer to step 3).

Prerequisites:
First, you need to install some libraries, such as numpy, pillow, pylab..
To install a library in windows:

Install PIP first, please refer to:
http://stackoverflow.com/questions/4750806/how-to-install-pip-on-windows

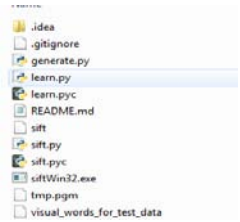
If you want to install a library called Pillow, you google 'python pillow windows' then you can get a link to python package index (https://pypi.python.org/pypi/Pillow/2.7.0).

The file format is like 'Pillow-2.7.0-cp26-none-win32.whl'. cp26 means the version is 2.6 and win32 means the pc is 32bits. Since my desktop is (Win7, 64bit, Python 2.7), I choose ''Pillow-2.7.0-cp27-none-win_amd64.whl'.

Then I use command 'python –m pip install Pillow-2.7.0-cp27-none-win_amd64.whl' to install Pillow.

After installing all the prerequisites, you can run the program now. I suggest all of you read the code first to understand how to extract visual key words.

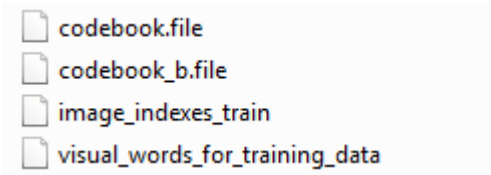
1. Please put 'siftWin32.exe' and 'sift' under your project folder. Now your folder like this.



2. Generate the codebook for a specific dataset with:
python learn.py –d path_to_folders_with_images
example:

```
D:\Minimal-Bag-of-Visual-Words-Image-Classifier-master>python learn.py -d D:\Ass
ignment1\ImageData\train\data\
```

After this step, you can get four files in the folder *path_to_folders_with_images*



codebook.file contains the sift features of all the visual words (we call it code book).
codebook_b.file is a binary version of codebook which will be used in step 3
image_indexes_train contains all the training image names
visual_words_for_training_data contains the visual words histogram for all the training images
I have trained the data already. You can download 4 files from the assignment folder.

3. To generate visual words histogram for a new test image use:
   python generate.py –c path_to_folders_with_images/codebook_b.file images_you_want_to_test

```
D:\Minimal-Bag-of-Visual-Words-Image-Classifier-master>python generate.py -c D:\
Assignment1\ImageData\train\data\codebook_b.file D:\Assignment1\ImageData\test\d
ata\bear\0018_375723120.jpg
--------------------
## extract Sift features
extracting Sift features
gathering sift features for D:\Assignment1\ImageData\test\data\bear\0018_3757231
20.jpg (183L, 128L)
--------------------
## loading codebook from D:\Assignment1\ImageData\train\data\codebook_b.file
--------------------
## computing visual word histograms
--------------------
## write the histograms to file
--------------------
```

   After this step you can the visual words histogram for a test image which is stored in the project folder called *image_indexe_test*.

The learning consists of:
1. Extracting local features of all the dataset images
2. Generating a codebook of visual words with clustering of the features
3. Aggregating the histograms of the visual words for each of the traning images

The generating consists of:
1. Extracting local features of the test image
2. Generating the histograms of the visual words for the test image using the prior generated codebook

How to call an exe from your java/python?
Please refer to the line37-43 in sift.py.