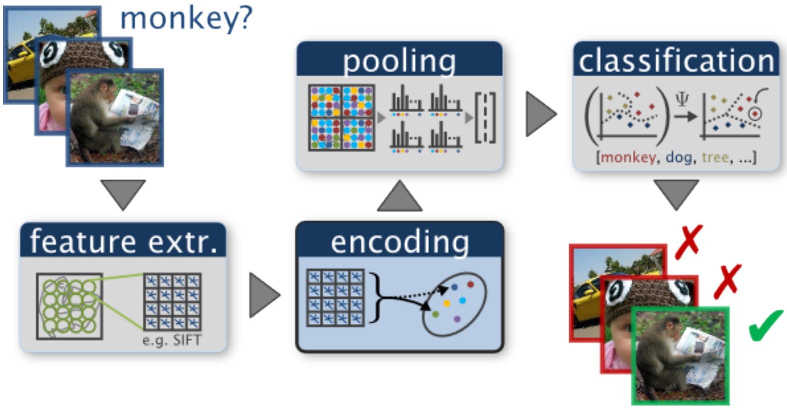


Bag of Visual Words

This is a README for how to use the files for classifying image datasets.

We will be needing a small mass-file renaming tool called GPRename (http://gprename.sourceforge.net/).

Unzip the **bagofwords.zip** and the following directories, along with the files will be created.

~/bagofwords
~/bagofwords/1.Vocabulary Builder/
~/bagofwords/1.Vocabulary Builder/CMakeLists.txt
~/bagofwords/1.Vocabulary Builder/vocabulary_builder.cpp
~/bagofwords/2.Histograms Calculator/
~/bagofwords/2.Histograms Calculator/Histograms/
~/bagofwords/2.Histograms Calculator/CMakeLists.txt
~/bagofwords/2.Histograms Calculator/histograms_calc.cpp
~/bagofwords/3.SVM trainer/
~/bagofwords/3.SVM trainer/CMakeLists.txt
~/bagofwords/3.SVM trainer/SVM_training.cpp
~/bagofwords/4.SVM tester/
~/bagofwords/4.SVM tester/File SVM tester/
~/bagofwords/4.SVM tester/File SVM tester/CMakeLists.txt
~/bagofwords/4.SVM tester/File SVM tester/SVM_test_file.cpp
~/bagofwords/4.SVM tester/Folder SVM tester/
~/bagofwords/4.SVM tester/Folder SVM tester/CMakeLists.txt
~/bagofwords/4.SVM tester/Folder SVM tester/SVM_test_folder.cpp

INSTRUCTIONS:

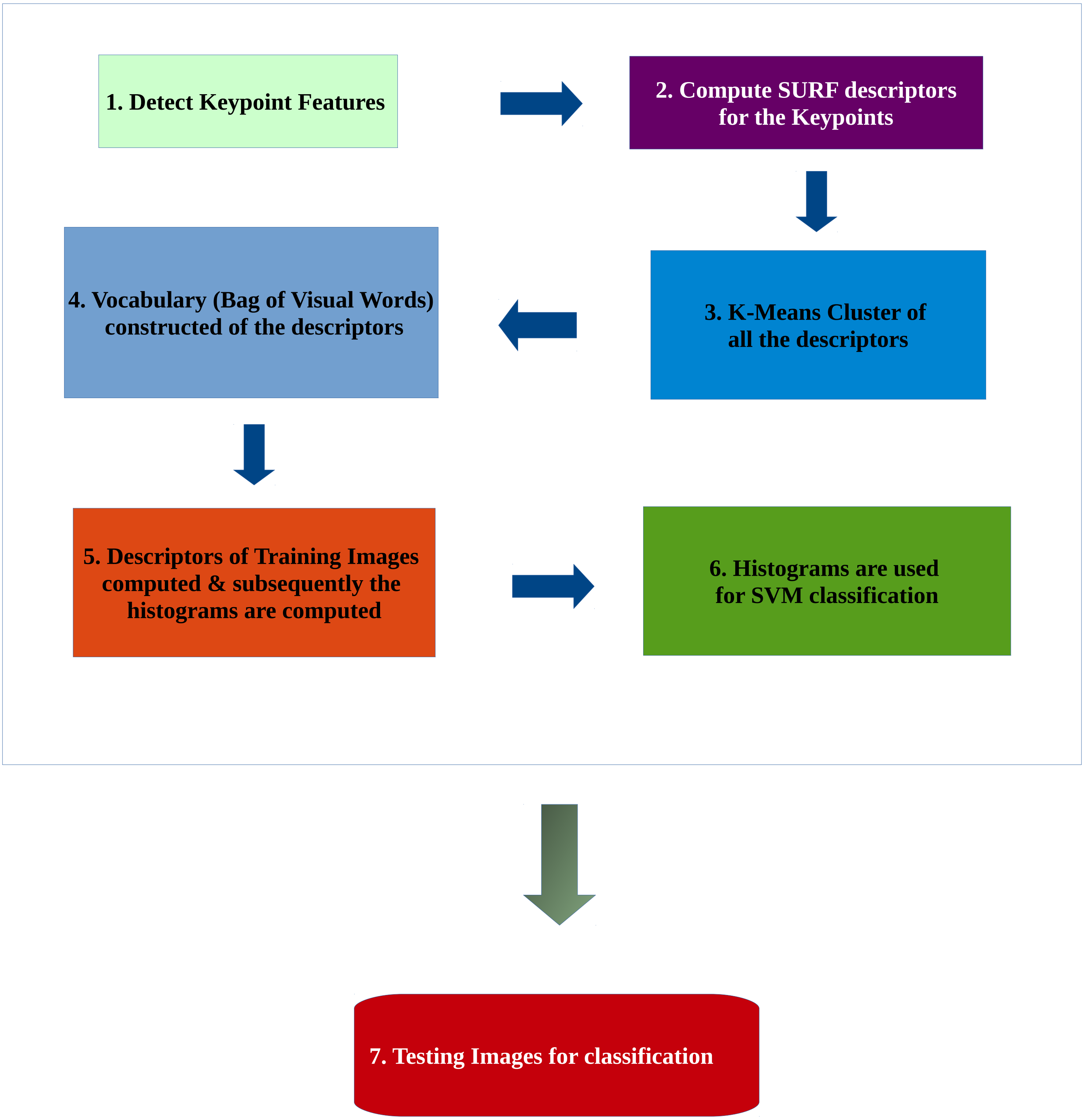
To start using the files , browse to the following directories and run the commands

```
>> cmake .
>> make
```

to compile and make the executable files

~/bagofwords/1.Vocabulary Builder/
~/bagofwords/2.Histograms Calculator/
~/bagofwords/3.SVM trainer/
~/bagofwords/4.SVM tester/File SVM tester/
~/bagofwords/4.SVM tester/Folder SVM tester/

PIPELINE



1. Detect Keypoint Features

2. Compute SURF descriptors for the Keypoints

3. K-Means Cluster of all the descriptors

4. Vocabulary (Bag of Visual Words) constructed of the descriptors

5. Descriptors of Training Images computed & subsequently the histograms are computed

6. Histograms are used for SVM classification

7. Testing Images for classification

1. Files and Folders

- *1. Vocabulary Builder*
    - vocabulary_builder.cpp
    - training_descriptors.yml (created by the Program)
    - vocabulary_color_1000.yml (created by the Program)
- *2. Histograms Calculator*
    - histograms_calc.cpp
    - SVM_training_data.yml (created by the Program)
    - /Histograms
- *3. SVM Trainer*
    - SVM_training.cpp
    - SVMdata.yml (created by the Program)
- *4. SVM Tester*
    - */File SVM tester*
        - SVM_test_file.cpp
    - */Folder SVM tester*
        - SVM_test_folder.cpp
    - classname.txt (created by the Program)

2. Usage

A) Vocabulary Builder :

The 1st directory " ***~/bagofwords/1.Vocabulary Builder/*** " has a file named **vocabulary_builder.cpp .**

Assuming it has been compiled using the makefile, the program can be run by running the following command in terminal with the required argument.

```
>> ./ vocabulary_builder [PATH TO IMAGE DATASET DIRECTORY]
```

Note :- The Image Dataset directory must be a folder with images in .jpg format and named in the pattern "0001.jpg","0002.jpg" and so on. For the purpose, the software GPRename can be used

vocabulary_builder.cpp does the following functions :-
1) Loads the image dataset directory
2) Detects the feature keypoints by using SURF detector
3) Computes the SURF descriptors for all the keypoints
4) Adds to a global matrix which stores all the descriptors of all the images to be used in building the vocabulary
5) KMeans clustering of the matrix to and the vocabulary(bag of words) is built

Exports the following files :-
1) **training_descriptors.yml** – the global matrix having all the descriptors (unclustered)
2) **vocabulary_color_1000.yml** – the bag of words built (clustered) to be used by the next program for building histograms

B) Histograms Calculator :

The 2nd directory “ **~bagofwords/2.Histograms Calculator/**” has a file name **histograms_calc.cpp .**

Assuming it has been compiled using the makefile, the program can be run by running the following command in terminal with the required argument.

```
>> ./ histograms_calc [PATH TO IMAGE DATASET DIRECTORY]
```

Note :- The Training Image Dataset directory must be a folder with images in .jpg format and named in the pattern “0001.jpg”,”0002.jpg” and so on. For the purpose, the software GPRename can be used

histograms_calc.cpp does the following functions :-
1) Loads the previously generated vocabulary
2) Creates a BOWImgDescriptorExtractor object based on FLANN matching and SURF descriptors and generates histograms for each training image in the path provided
3) Exports the histograms of each image to the folder “./Histograms”
4) Builds a matrix with all the histograms stacked row-wise, to be used for SVM classification purposes

Exports the following files : -

1) /**Histograms/*.yml** – individual histograms
2) **SVM_training_data.yml** – histograms of all the training images stacked row-wise to be used by the SVM trainer program for classification

C) SVM trainer :

The 3rd directory “ **~bagofwords/3.SVM trainer/**” has a file name **SVM_training.cpp .**

Assuming it has been compiled using the makefile, the program can be run by running the following command in terminal with the required argument.

```
>> ./ SVM_training
```

The **SVM_training.cpp** file does the following functions :-
1) Reads in the SVM training data (histograms)
2) Asks for the label names for the classes
3) Asks for the number of objects to be classified at class 1
4) Exports the SVM data to be used for SVM tester
5) Exports the class names to be used for SVM tester

Exports the following files : -

1) **SVMdata.yml** – SVM data exported to be used for SVM testing
2) **../4.SVM tester/classname.txt –** holds the classnames which were given as input , to be used by the SVM tester in displaying the results

D) SVM tester :

The 2nd directory " **~bagofwords/4.SVM tester/**" has 2 folders **File SVM tester** and **Folder SVM tester.**

1) **File SVM tester** – This program will test the class for a single test image. This directory has a file **SVM_test_file.cpp**

Usage :

```
./SVM_test_file [TEST IMAGE FILEPATH]
```

2) **Folder SVM tester** – This program will test the class for all the test images in the directory. This directory has a file **SVM_test_folder.cpp**

Usage :

```
./SVM_test_folder [TEST IMAGE DATASET FOLDER PATH]
```

Note :- The Test Image Dataset directory must be a folder with images in .jpg format and named in the pattern "0001.jpg","0002.jpg" and so on. For the purpose, the software GPRename can be used

The programs do the following functions :

1) Read the SVM training data
2) Read the vocabulary
3) Load the test image(s) and compute the histograms by BOWImgDescriptorExtractor against the vocabulary
4) Reads the class names from the classname.txt
5) SVM predicts the class to which the image belongs

Exports :

Asks the user whether or not to export the data