1 Quantification

The Quantification is the step we will use for determine which images correspond to which species. For that we will use the bag of words in a first place to simplify the signature of whole images. And after use the K-nn method for the classification.

1.1 Bag of words

In the bag of words method we found two different step, the calculation of K-means and the design of a signature for the images.

1.1.1 K-means

The K-means is a simply method which consist to reduce the number of points or vectors in our case. The first step is to determinate randomly k centroid vectors, after with an Euclidean distance (1) we attribute the descriptors to whole images to the nearest centroid vectors.

$$\sum_{k=0}^{centroid} \sum_{i=0}^{desc} \parallel x_k - u_i \parallel^2 \tag{1}$$

The last stage is an update step, for each centroid vectors we calculate the means of whole the descriptors associate to, so we obtain a new centroid vectors. And we applicate this algorithm for x iteration choose by the user.

For an application on a cloud of points we obtain this kind of result:

1.1.2 Signature

The creation of the signature is the last step of the bag of words method, they consist to assign to each images a signature in function of the number of word in the image.
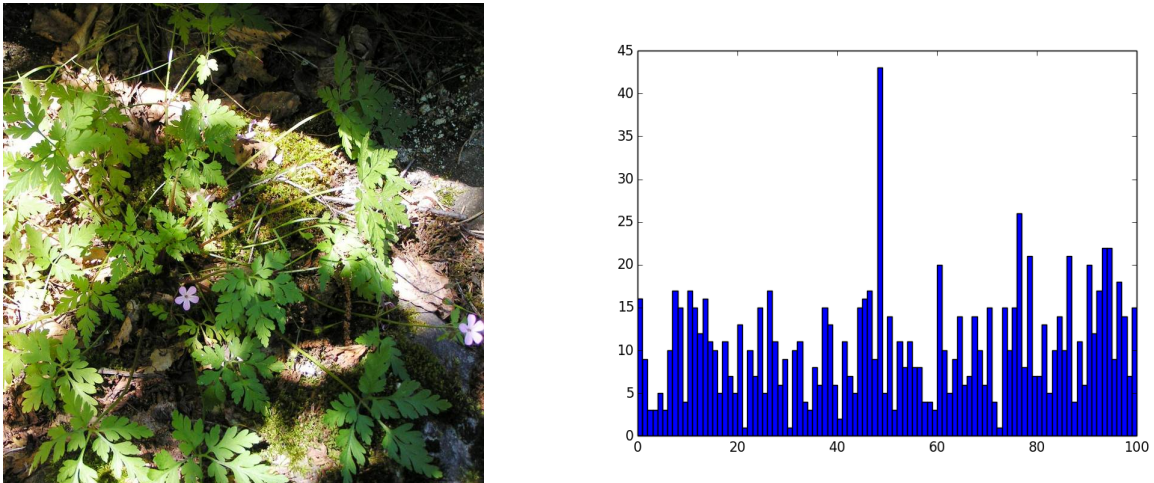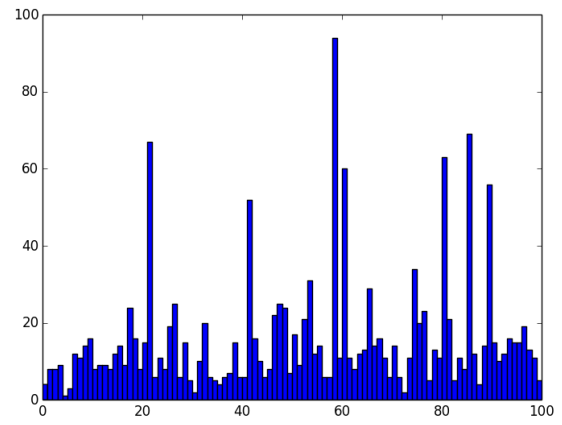

Figure 1: Signature de l'image 128

Figure 2: Signature de l'image 132

1.2   K-nn