Ecole Supérieure Informatique Sidi Bel Abbes

Second Cycle

Computer Vision

LAb -3-: Bag of Visual Words(BoVW) for Image Classification

October 26, 2025

1. **Objective:**

   - BoVW is a commonly used technique in image classification. The idea behind this technique, is similar to the bag of words in NLP but in this technique we use image features as words.

2. **Programming Language and Library**

   - UC Merced Land Use Dataset.
   - Python with OpenCV and scikit-learn libraries.

3. **Lab Procedure:**

   - Data Loading and Preprocessing (15 minutes):
   - Dataset Splitting:
     - Split the dataset into a training set and a testing set. A common split ratio is 80% for training and 20% for testing.
     - Ensure that the split maintains a balance of images from different categories.
   - **SIFT Feature Extraction :** For each image in the dataset:
     - Detect and compute SIFT key points and descriptors,
     - Store the descriptors along with their corresponding class labels.

```python
import numpy as np
import cv2 as cv
img = cv2.imread('images/dolphin.jpg')
crop_img = img[200:, 200:]
gray= cv.cvtColor(img,cv.COLOR_BGR2GRAY)
sift = cv.SIFT_create()
#kp = sift.detect(gray,None)
kp, des = sift.detectAndCompute(gray,None)
img=cv.drawKeypoints(gray,kp,img)
cv.imwrite('sift_keypoints.jpg',img)
plt.imshow(img)
plt.show()
```

   - Building a Visual Vocabulary:
     - Cluster the extracted features from all training images into a set of visual words using a clustering algorithm such as K-means. Each cluster center becomes a visual word.
   - Quantization:

  - For each image, assign local feature descriptors to the nearest visual word (cluster center). This step is known as quantization.
- Histogram Representation:
  - Represent each image as a histogram of visual words. Count the occurrences of each visual word in the image.
- Normalization:
  - Normalize the histograms to account for variations in the number of features in each image. Common normalization methods include L1 or L2 normalization.
  - L1 Normalization: Ensures that the sum of all histogram values is 1, giving a measure of frequency for each bin.
    L1 normalized histogram = $\frac{h}{\sum |h|}$]
  - L2 Normalization: Scales the histogram such that the sum of squares of the values equals 1, often making the features invariant to scale differences. L2 normalized histogram = $\frac{h}{\sqrt{\sum h^2}}$]
- Classifier Training:
  - Train a classifier (Support Vector Machine) using the normalized histograms as input features and the corresponding class labels.
- Generate Predictions:
  - For each test image, extract features, create its histogram representation using the visual vocabulary, and use the trained classifier to predict the class.
- Performance Evaluation:
  - Compute accuracy to evaluate the SVM classifier's performance.
  - Draw and include the confusion matrix in your lab report.

4. **Assignment:** Provide a report on the outcomes.

.