



Due Date: 23:59 pm on Sunday, April 14th, 2024

## Image classification using Bag of Visual Words

In this assignment, you will classify the images using Bag of Visual Words (BOVW) using keypoint description methods (SIFT/SURF and ORB) and KMeans clustering method. You will also compare the methods (SIFT/SURF and ORB) with respect to run time and accuracy. First of all, you will extract and obtain the multiple keypoints from sub-images by using an keypoint description method. Then you will compare and match these keypoints to classify the images (See figure 1). You will implement BOVW framework by yourselves. You will use subset of the given dataset [3].

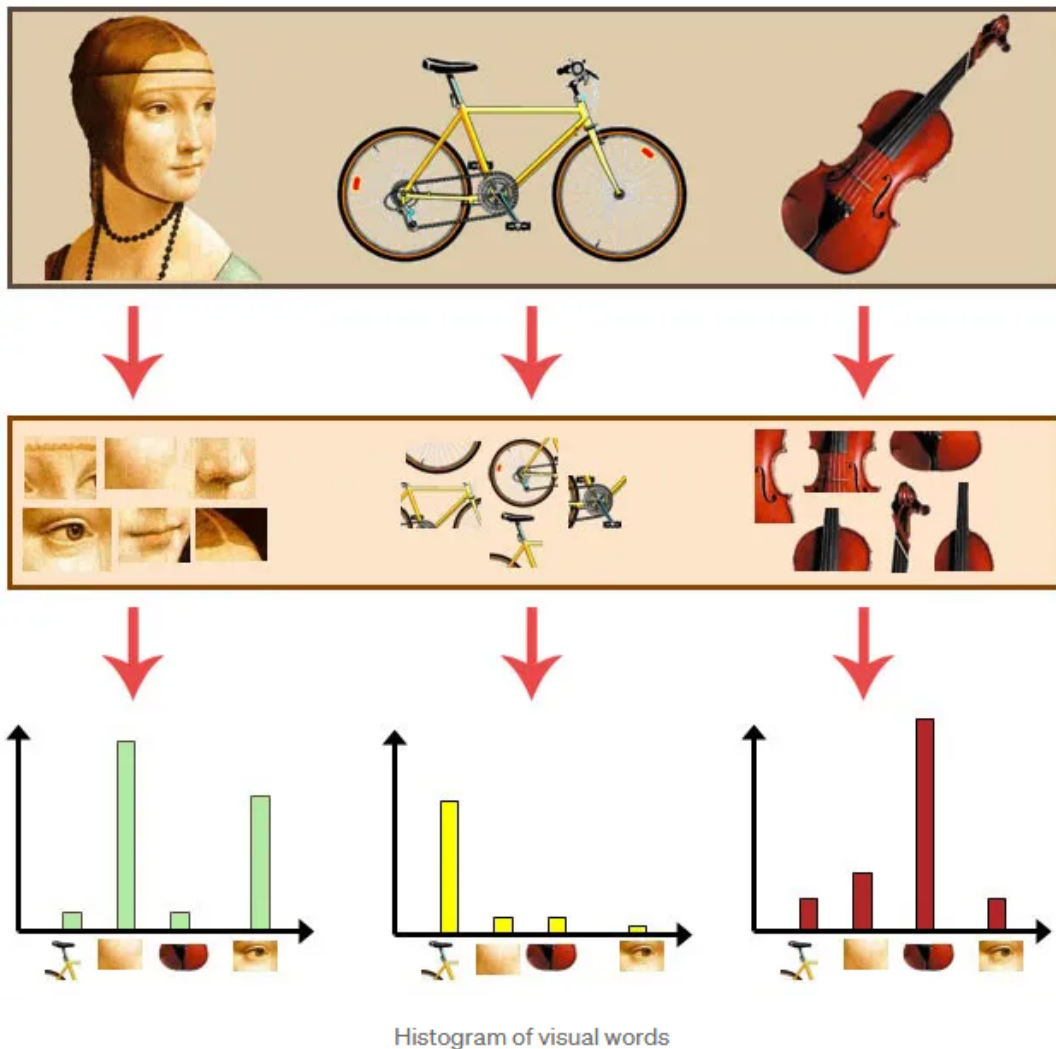


Figure 1: Visual of Procedure to Be Used for Image classification [1]

The Bag of Visual Words (BOVW) concept involves representing an image as a collection of features, which include keypoints and descriptors. Keypoints are distinctive points within an image that remain consistent even when the image is transformed (e.g., rotated, resized). Descriptors are descriptions of these keypoints. By utilizing keypoints and descriptors, vocabularies are constructed, and each image is represented as a frequency histogram of features

present in the image. This histogram can then be used to find similar images or predict the category of the image.

To implement the process, features are first detected, and descriptors are extracted from each image within the dataset. Subsequently, a visual dictionary is constructed. Feature detection and descriptor extraction are achieved through the utilization of feature extractor algorithms. Following descriptor extraction, clusters are formed using the descriptors. This clustering process can employ algorithms like K-Means, DBSCAN, or others. The centroids of these clusters serve as the visual dictionary's vocabularies. In the final step, a frequency histogram is created for each image using the visual dictionary's vocabularies and their occurrences within the image. These histograms constitute the Bag of Visual Words (BOVW) representation for the respective images.

## The Implementation Details

1. **Keypoint detection (10 Points):** Firstly you are expected to use SIFT or Harris-Laplacian keypoint detection.
2. **Feature Extraction (10 Points):** Then, you are expected to extract keypoints in the images by a keypoint extraction method (SIFT/SURF and ORB). You can use built-in methods (`cv2.SIFT_create`, `cv2.xfeatures2d.SURF_create`, `cv2.ORB_create`).
3. **Feature Matching (10 Points):** Then you are expected to implement feature matching technique based on Euclidean distance.
4. **BoW formation (20 Points):** Then you should Cluster the features, form the BoW dictionary and quantize images.
5. **Classification (50 Points):** You should classify images using k-NN approach and evaluate performance of different visual vocabularies. You can use built-in methods (`sklearn.neighbors.KNeighborsClassifier`).
  - (a) Implementation of k-nn approach using different distance measures (such as Cosine similarity, Chi-square distance) (10pts).
  - (b) Evaluation of different techniques (different BoW histogram sizes, different distance measures, SIFT vs ORB) in terms of run time, classification accuracy, confusion matrix (30pts).
  - (c) Discussion of the results (10 pts)
6. You should pay attention to code readability such as comments, function/variable names and your code quality:
  - 1) no hard-coding
  - 2) no repeated code
  - 3) cleanly separate and organize your code
  - 4) use consistent style, indentation.
7. You should use Python 3 for the assignment.

## What should you write in the notebook?

In your notebook, firstly, you are expected to write a detailed explanation (coding details, about your calculations and implementation details) for the five main steps:

- Keypoint detection,
- Feature Extraction,
- Feature Matching,
- BoW formation,
- Classification.

Secondly, you are expected to present the following results:

- Plots showing feature points for some example images,
- Table for runtime and visual comparison of the description methods (SIFT/SURF and ORB).
- Related confusion matrices and comparison tables as discussed above.

Finally, you are expected to comment about your results (Explanation about your intermediate results and commenting about your program's performance of your classified images).

## What to Hand In

Your submission format will be:

- `b<studentNumber>.ipynb`

Archive this folder as `b<studentNumber>.zip` and submit to `https://submit.cs.hacettepe.edu.tr`.

## Grading

The assignment will be graded out of 100 according to the given grade distribution as follows:

- For each part, 0% (no implementation), 20% (an extremely incomplete implementation), 40% (an incomplete implementation), 60% (a partially correct implementation), 100% (a correct implementation).

## Academic Integrity

All work on assignments must be done individually unless stated otherwise. You are encouraged to discuss with your classmates about the given assignments, but these discussions should be carried out in an abstract way. That is, discussions related to a particular solution to a specific problem (either in actual code or in the pseudocode) will not be tolerated. In short, turning in someone else's work, in whole or in part, as your own will be considered as a violation of academic integrity. Please note that the former condition also holds for the material found on the web as everything on the web has been written by someone else.

## References

[1] <https://towardsdatascience.com/bag-of-visual-words-in-a-nutshell-9ceea97ce0fb>

[2] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1-10, 2018.

[3] <https://github.com/fastai/imagenette?tab=readme-ov-file>