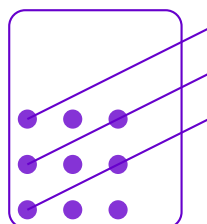


Bag of Visual Words for Image Classification

A Computer Vision Project using SIFT, KMeans, KNN & SVM

Muhammad Humaiyon

July 10, 2025



1 Overview

This project explores a Bag of Visual Words (BoVW) pipeline for image classification using handcrafted SIFT descriptors, KMeans clustering, and KNN/SVM classifiers. It aims to classify grayscale images of three object categories: **cats**, **chairs**, and **cameras**.

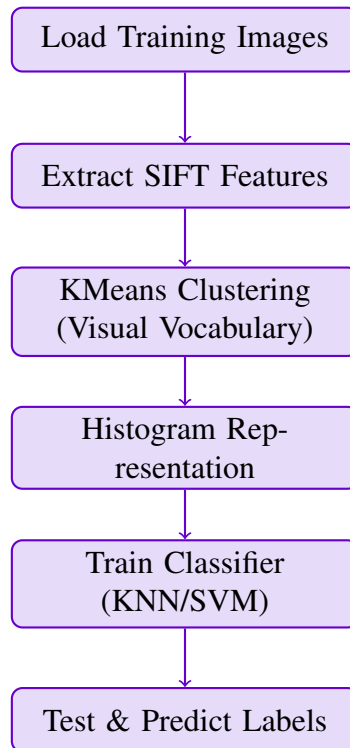
2 Objective

- Build a robust feature extraction pipeline using SIFT.
- Apply KMeans to create a visual vocabulary.
- Train KNN and SVM models to classify unseen test images.
- Visualize histogram representations and predictions.

3 Tools and Technologies

-  Python with OpenCV
-  SIFT for feature extraction
-  KMeans Clustering
-  KNN & SVM Classifiers
-  Custom Image Dataset

4 Methodology Flow



5 Feature Extraction

Each image is converted to grayscale. SIFT keypoints and descriptors are extracted. Descriptors are stacked across all images to form a feature pool.

6 Visual Vocabulary (KMeans)

- All SIFT descriptors are clustered using KMeans ($k = 200$).
- Each cluster center becomes a **visual word**.
- Each image is encoded as a histogram of these words.

7 Classification using KNN

- Histograms from training images are used as input.
- A K-Nearest Neighbor classifier ($k=5$) is trained.
- Tested on final cat and chair images.

8 Classification using SVM

- Three-class setup: Cat, Chair, Camera
- Linear SVM classifier trained on BoVW histograms.
- Predictions are made on unseen test images.

9 Visualizations

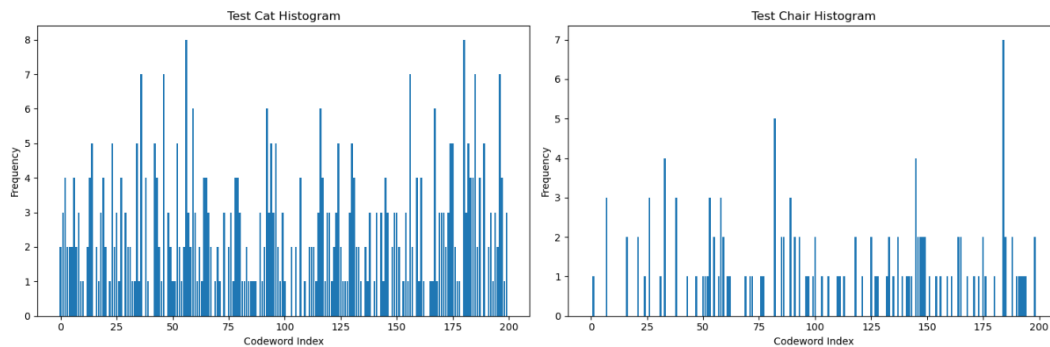


Figure 1: Visual Word Histogram for a Test Image



Figure 2: Predicted Class Labels for Test Images (Cat, Chair, Camera)

10 Key Learnings

- BoVW works effectively with SIFT and simple classifiers.
- Histogram encoding reduces image complexity drastically.
- Visualizing results helps debug class confusion.

“This project taught me how to represent images using bag-of-words techniques, and how clustering + classical ML can be combined to build powerful vision systems.”