CSE463 HW2

Ertugrul Kasikci 200104004097

May 30, 2024

1 Introduction

In this homework, a simple object recognition algorithm based on feature detectors and descriptors from the OpenCV library was implemented. Ten objects were selected from the turntable dataset, and three different feature detectors and descriptors were used to learn and recognize these objects.

2 Object Selection

The following objects were selected for this experiment:

1. toothbrush_4
2. shampoo_6
3. onion_1
4. marker_6
5. hand_towel_5
6. garlic_4
7. food_box_11
8. cereal_box_5
9. calculator_1
10. bell_pepper_2

3 Feature Detectors and Descriptors

- ORB (Oriented FAST and Rotated BRIEF)
- BRISK (Binary Robust Invariant Scalable Keypoints)
- AKAZE (Accelerated-KAZE)

4 Algorithm Details

4.1 Data Preparation

The turntable dataset is downloaded and the images for the 10 chosen objects are selected. Each object's images were divided into training and testing sets with a 90/10 split. The training set was balanced to ensure each class had the same number of images.

4.2 Feature Extraction

For each method, features using the respective detectors and descriptors were extracted. The detected keypoints and computed descriptors were used to form the feature vectors for each image.

4.2.1 Descriptor Statistics

Statistics about the descriptors for each method are provided below:

- **ORB**:
  - Number of images: 2475
  - Average number of descriptors per image: 128.31
  - Descriptor dimensions: 32
- **BRISK**:
  - Number of images: 3256
  - Average number of descriptors per image: 89.80
  - Descriptor dimensions: 64

- **AKAZE**:
  - Number of images: 2326
  - Average number of descriptors per image: 35.01
  - Descriptor dimensions: 61

The number of descriptors per image indicates how many features were detected in each image on average. Descriptor dimensions refer to the length of the feature vectors used to represent each keypoint.

4.3 Training

K-means clustering was used to cluster the descriptors into 4096 clusters. The cluster labels were then mapped to the class labels based on the most common true label in each cluster.

4.4 Prediction

For the test images, features were extracted and assigned to clusters using the trained k-means model. The predicted class for each image was determined by the majority vote of the predicted cluster labels.

5 Keypoint Detection Visualizations

For each object and each feature detector, keypoints were detected and visualized. All the visualizations are saved in the 'output' folder inside the project directory. Below is an example for the object 'calculator_1':

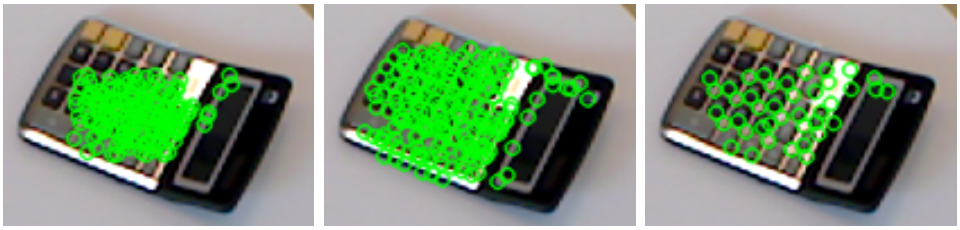

Figure 1: Keypoint detection for calculator_1 using ORB, BRISK, and AKAZE (from left to right)

6 Results

The performance of the object recognition algorithm was evaluated using accuracy and confusion matrices for each feature detection method.

6.1 ORB

- **Accuracy:** 0.5045
- **Confusion Matrix:**

	toothbrush_4	shampoo_6	onion_1	marker_6	hand_towel_5	garlic_4	food_box_11	cereal_box_5	calculator_1	bell_pepper_2
toothbrush_4	2	2	4	12	0	0	0	0	0	0
shampoo_6	0	21	32	3	0	0	0	0	0	0
onion_1	0	0	54	1	0	0	0	0	0	0
marker_6	0	0	1	78	0	0	0	0	0	0
hand_towel_5	0	0	0	0	0	0	0	0	0	0
garlic_4	0	6	15	1	0	0	0	0	0	0
food_box_11	0	0	20	3	0	0	9	0	0	0
cereal_box_5	0	0	5	1	0	0	0	0	0	0
calculator_1	0	2	39	5	0	0	0	0	0	0
bell_pepper_2	0	0	14	1	0	0	0	0	0	6

6.2 BRISK

- **Accuracy:** 0.6523
- **Confusion Matrix:**

	toothbrush_4	shampoo_6	onion_1	marker_6	hand_towel_5	garlic_4	food_box_11	cereal_box_5	calculator_1	bell_pepper_2
toothbrush_4	0	3	8	12	0	0	2	0	0	0
shampoo_6	0	58	0	0	0	0	0	0	0	0
onion_1	0	0	54	1	0	0	0	0	0	0
marker_6	0	0	0	79	0	0	0	0	0	0
hand_towel_5	0	1	4	1	0	0	0	0	0	2
garlic_4	0	23	16	8	0	9	0	0	0	1
food_box_11	0	0	10	2	0	0	55	0	1	0
cereal_box_5	0	1	2	3	0	0	0	0	0	0
calculator_1	0	16	34	3	1	0	0	0	11	0
bell_pepper_2	0	1	1	1	0	0	0	0	0	36

6.3 AKAZE

- **Accuracy:** 0.8500
- **Confusion Matrix:**

	toothbrush_4	shampoo_6	onion_1	marker_6	hand_towel_5	garlic_4	food_box_11	cereal_box_5	calculator_1	bell_pepper_2
toothbrush_4	5	1	2	1	0	0	0	0	0	0
shampoo_6	0	54	1	3	0	0	0	0	0	0
onion_1	0	0	54	0	0	0	0	0	0	0
marker_6	0	0	0	79	0	0	0	0	0	0
hand_towel_5	0	0	0	0	0	0	0	0	0	0
garlic_4	0	4	4	3	0	1	0	0	0	0
food_box_11	0	1	2	1	0	0	21	0	0	1
cereal_box_5	0	0	0	0	0	0	0	0	0	0
calculator_1	0	6	9	1	0	0	0	0	22	0
bell_pepper_2	0	2	2	0	0	0	0	0	1	19

7 Discussion

The results indicate that the AKAZE method performed the best, with an accuracy of 0.8500, followed by BRISK and ORB. The confusion matrices show varying degrees of misclassification among the objects. The performance differences can be attributed to the nature of the feature detectors and descriptors used. The reasons for the failures could include:

- Insufficient distinct features in some objects.
- Similar appearance of certain objects, leading to misclassification.
- The limitations of the feature detectors and descriptors in capturing unique object characteristics.