Product Recognition on Store Shelves
IPCV Project WS23/24

Freddy Fernandes
freddy.fernandes@studio.unibo.it

1. Introduction

For this project, the objective is to detect objects in supermarket scenarios by recognising certain products on store shelves. We focus specifically on identifying and recognising cereal boxes from various brands on store shelves.

2. Task A

In the first stage of our project, we aim to identify a single instance of each product in a picture. We have images of the products to identify in a set of images called 'Models' and a set of images of supermarket shelves called 'Scenes'. We use a method that begins by identifying key points and descriptors in both the product's image and the image of the store shelf using a technique called Scale-invariant feature transform (SIFT). We then match these key points from the product to those found in the shelf image using a function called FeatureMatch, which relies on a KNN FLANN-based matcher. If we find enough matching points, we use RANSAC to create a homography. We ensure that our outline is accurate by checking if it's the right shape for a cereal box and if its colours match the product, ensuring we've correctly identified the product.

To correctly identify similar-looking products with similar designs, we need to consider the colour of the model images too. We split our images into bins and computed mean intensities in each bin for all colour channels (r,g,b). Then, we compare these average colours between the corresponding bins of the model and target images. If the colour difference in any bin exceeds a certain threshold, we mark it as invalid. If too many bins are invalid, we discard the bounding box. Should a bounding box not pass the colour filtering, it is blacked out in the target image to prevent its future detection, thus addressing potential false positives due to colour invariance in SIFT.

The project also consists of functions like display_model_images to provide a simple representation of the models being considered and Output to visually present the models and the detection results, respectively. For each detected product, the system provides comprehensive information, including the name of the product, the number of instances found, their dimensions, and their positions within the scene.

Results:
The system consistently is able to detect all the products correctly in this task. However, altering the parameters like colour threshold and min number of matches slightly can lead to some true positives being missed (eg model 11 in scene 0)

Model 0: Nesquik

Model 1: CK blue

Model 11: CK orange

Model 19: Country crisp

Model 24: Fitness

Model 26: Nesquik Duo

Model 25: CocoPops


```
Product Nesquik (Model 0) - 1 instance/s found:
        Instance position: (161, 212), width: 311px, height: 437px
Product CK blue (Model 1) - 0 instance/s found:
Product CK orange (Model 11) - 1 instance/s found:
        Instance position: (443, 179), width: 299px, height: 391px
Product Country crisp (Model 19) - 0 instance/s found:
Product Fitness (Model 24) - 0 instance/s found:
Product Nesquik Duo (Model 26) - 0 instance/s found:
Product CocoPops (Model 25) - 0 instance/s found:
```


```
Product Nesquik (Model 0) - 1 instance/s found:
        Instance position: (170, 233), width: 325px, height: 444px
Product CK blue (Model 1) - 1 instance/s found:
        Instance position: (819, 194), width: 301px, height: 412px
Product CK orange (Model 11) - 1 instance/s found:
        Instance position: (478, 192), width: 295px, height: 400px
Product Country crisp (Model 19) - 0 instance/s found:
Product Fitness (Model 24) - 0 instance/s found:
Product Nesquik Duo (Model 26) - 0 instance/s found:
Product CocoPops (Model 25) - 0 instance/s found:
```

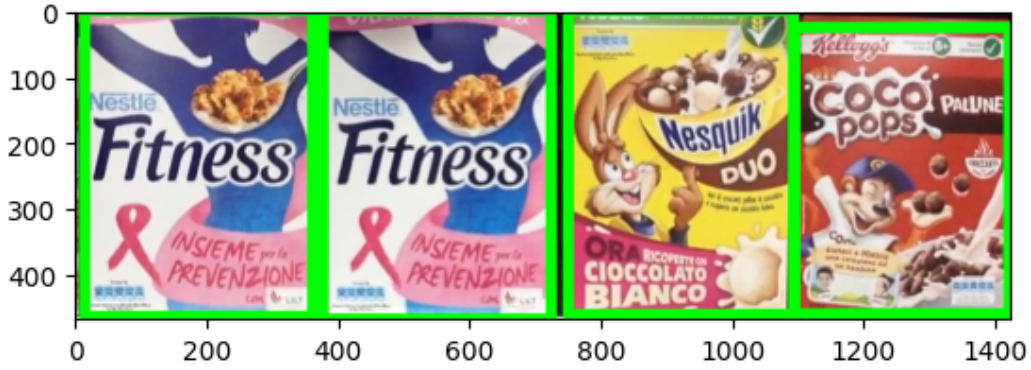
3. Task B

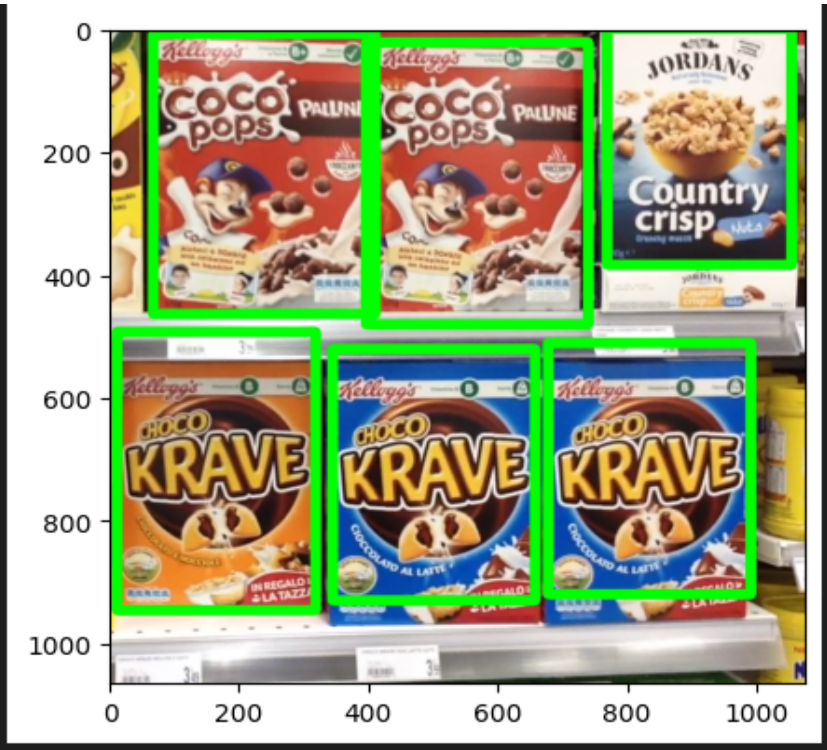
In the second task of this project, the goal is to extend the detection capabilities to identify one or more instances of various products within a given scene. This task requires a more complex approach, combining local invariant features with the Generalized Hough Transform (GHT) for better accuracy. This method is applied to the model-target image.

The process begins similarly to Step A, with the extraction of SIFT keypoints and descriptors. The system then uses the KNN Flann Matcher for feature matching. The new step here is using GHT (Generalized Hough Transform) for better accuracy. Here, the system computes joining vectors from the model's centre point (the reference point) and resizes these vectors based on the scale of matching keypoints. Votes are added into an Accumulator Array for every scaled joining vector, allowing the system to detect peaks that correspond to potential product locations in the scene.

Next, the system draws bounding boxes around these possible product locations using the centre point as a reference. The size of each box is determined by using a scaling factor that comes from the average size of the matched features. To make sure we're identifying products correctly, the system checks the size and colour of these boxes, similar to what we did in Step A. Plus, if any boxes are detected too close to each other, they are combined into one and their coordinates are the mean between the merged boxes.
In the end, we use the Output function again to show the detection results in the specified scene images. This time the output lists out details for the multiple instances of the models found in each scene image.

Results:
The system is consistently able to detect all the products correctly in this task too. There are no instances of missed/incorrect detections.






```
Product Nesquik (Model 0) - 0 instance/s found:
Product CK blue (Model 1) - 2 instance/s found:
        Instance 1 position: (500, 725), width: 313.0px, height: 409.0px
        Instance 2 position: (833, 715), width: 313.0px, height: 409.0px
Product CK orange (Model 11) - 1 instance/s found:
        Instance 1 position: (164, 719), width: 307.0px, height: 453.0px
Product Country crisp (Model 19) - 1 instance/s found:
        Instance 1 position: (910, 192), width: 285.0px, height: 384.0px
Product Fitness (Model 24) - 0 instance/s found:
Product CocoPops (Model 25) - 2 instance/s found:
        Instance 1 position: (238, 236), width: 343.0px, height: 456.0px
        Instance 2 position: (567, 251), width: 343.0px, height: 459.0px
Product Nesquik Duo (Model 26) - 0 instance/s found:
```

4. References:

- OpenCV Documentation
- Feature extraction and image classification Project
- Generalized Hough Transform Tutorial
- Hough Transform Guide
- Finding Product Images in Shelf Images Project
- Product_Recognition_Project
- Product_Recognition_Project
- Product-Recognition-on-Store-Shelves Project
- Hough Transform in OpenCV
- Computer Processing and Image Processing Lab 5 Unibo