COMPUTER VISION

Object detection and Feature Matching

Middle course test - April 2025

Objective

The goal of this project is to develop an object detector capable of locating known objects in an input image. Three categories are considered: “power drill”, “mustard bottle”, “sugar box”. The system should detect the position of each known object in the image and highlight it by means of a bounding box, see example below.

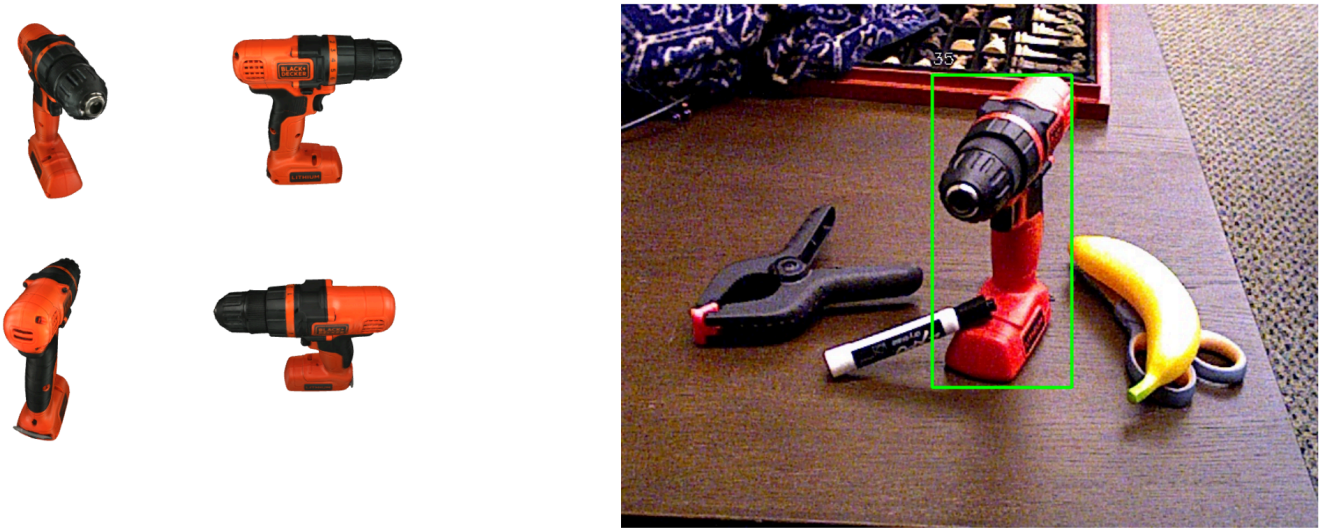

Figure 1: Example of the system output. On the left, the set of templates composed of different synthetic views of the object of interest to be localized. On the right, an input image of a scene with a bounding box highlighting the detected object.

Project Description

The system is organized as follows.
The **inputs** are:

- A RGB image of a scene;
- A set of synthetic views for each object to be recognized;
- A set of binary masks for each synthetic view.

Each mask distinguishes the object (pixel value 255) from the background (pixel value 0) – the mask is part of the dataset and is made available, but it might be neglected from your system.

The **outputs** of the system are:

- a set of bounding box coordinates of each detected object, saved to a text file;
- an image showing the bounding boxes superimposed onto the original image.

It is suggested to develop your system based on a robust feature matching, but alternative approaches (appearance based, template matching, bag of words) are also accepted.

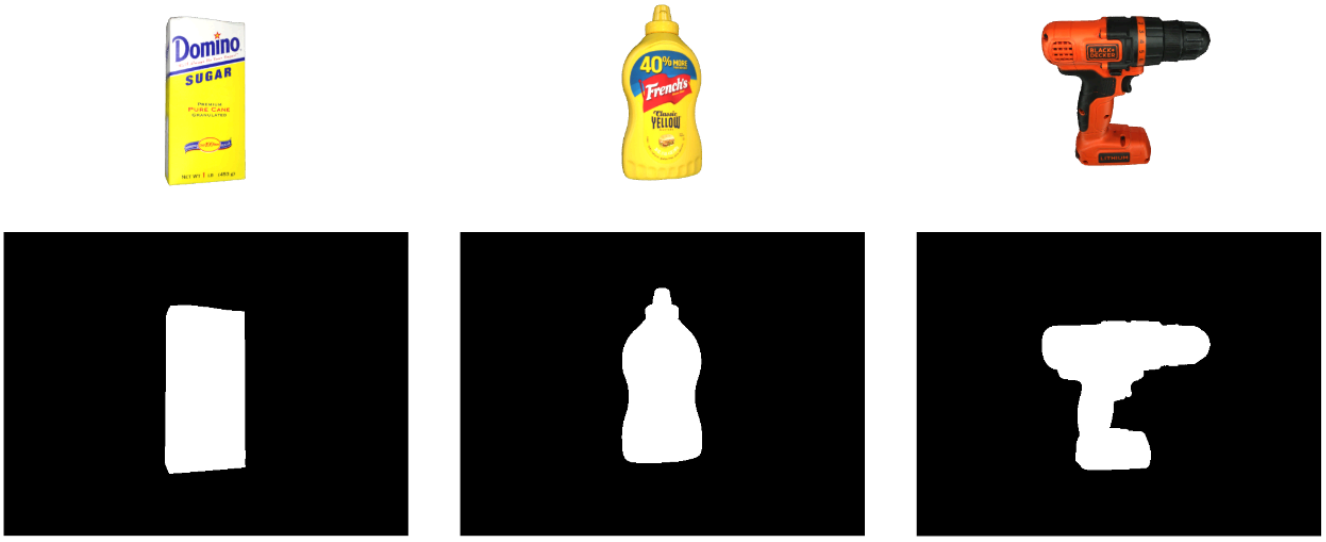

Figure 2: The set of object items to be recognized by the system. From the left to the right: a sugar box, a mustard bottle, a power drill. For each object item a set of synthetic views is provided, both color images and binary segmentation masks.

Dataset Description and Performance Measurement

The dataset to be used for the development of the project is available at this link:

https://drive.google.com/drive/folders/1heXAbX4WKXf3-z2sl68Qg-cvbcVwosxO?usp=sharing

The dataset[1] considers three objects. For each one, the following folders are provided:

- test_images, containing the images on which the developed system should be tested;
- models, containing 60 views of the object and the corresponding binary segmentation masks;
- labels (AKA ground truth), containing annotations for each image and evaluating the position of the objects to be located.

Annotations are provided as a text file (.txt), having one line for each object of interest in the image. Each line is a string including id and object name, and the coordinates of the bounding box vertices. For example:

```
<object_id>_<object_name> <xmin> <ymin> <xmax> <ymax>
```

[1] The provided dataset has been extracted from the public available YCB Video dataset commonly used for 6D object pose estimation (https://paperswithcode.com/dataset/ycb-video).

Evaluation images contain the object seen from various viewpoints, with changes in illumination, occlusions and in presence of objects of a different category. The system to be developed must be robust to all these situations and the robustness of the developed system to these variations must be analyzed in the final report.

To assess the performance of your object detection pipeline, the following metrics should be computed using the evaluation images and ground truth information provided in the dataset: Mean Intersection over Union (mIoU)[2] and Detection accuracy. Such metrics shall be used to evaluate the system performance as follows:

- The mean Intersection over Union (mIoU) is the average of the IoU computed for each object category (sugar box, mustard bottle, power drill);
- For detection accuracy, the number of object instances correctly recognized for each object category, considering a true positive (object is correctly detected) if the predicted and ground truth bounding boxes have IoU>0.5).

Project delivery

The project must be developed in C++ with the OpenCV library **only**. **The project cannot be developed using deep learning**.

You need to deliver your project including:

- All the source code (C++), where each source file must contain the name of the main group member developer – **one author per file is allowed**; you should check that **the code compiles on the Virtual Lab**, that is considered the official building environment;
- A report (maximum two pages of text excluding images and tables) presenting your approach and **the performance measurement on the dataset provided and linked above**. You shall report **the metrics and the output images** for every element in the dataset.

When delivering your project, you should **clearly identify** the contribution of each member in terms of ideas, implementation, tests and performance measurement. You can organize the work as you prefer: you are not forced to assign one specific step to each group member. Please also include **the number of working hours** per person in the report. This is needed for a monitoring on our side of the effort requested – the evaluation will not depend on the number of working hours.

It is not allowed to include code that was not written by any of the group members - this includes ChatGPT & friends.

[2] https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/