

# Image classification with SVM

Andrea Portscher, Juan Pablo Stumpf, Philip Wille

January 20, 2021

## Introduction

For this project we implemented an image classifier using three different feature extraction approaches: Scale-invariant feature transform (SIFT), Speeded up robust features (SURF) and Histogram of Gradients (HOG). First, we prepared an image dataset containing five different image classes, each comprising training and testing data. After extracting features using the above algorithms, we trained a Support-Vector Machine (SVM) and tested, how well it was able to classify images.

## Approaches

In the following section we will shortly explain the functionality of the feature extraction algorithms we applied to the dataset - namely SIFT, SURF and HOG. In general, feature extraction algorithms work in three steps[1]:

1. They select interest points at distinctive locations, such as corners.
2. The neighbourhood of the interest points is represented by a feature vector.
3. The descriptor vectors are matched between different images.

### SIFT

### SURF

### HOG

HOG stands for Histogram of Oriented Gradients and is a feature descriptor used in computer vision and image processing. Its purpose is to count the occurrence of gradient orientations in a given image. The first time HOG was described by Robert K. McConnell of Wayland Research Inc. in a patent application in 1986 without using the term. It has become more famous when Navneet Dalal and Bill Triggs presented their work on HOG descriptors in their 2005 published paper.

## Functionality of HOG

HOG decomposes an image into small squared cells, computes a histogram of oriented gradients in each cell, normalizes the result using a block-wise pattern, and return a descriptor for each cell. Stacking the cells into a squared image region can be used as an image window descriptor for object detection, for example by means of an SVM.

## Preprocessing

For using the HOG there are some recommended preparation steps for the provided data. Firstly, it's recommended to resize the image to a size of 64 pixels by 128 pixels for receiving the best results. This is because the standard implementation of the HOG, described in the above mentioned paper of Navneet Dalal and Bill Triggs, divides the image into  $8 * 8$  and  $16 * 16$  patches for extracting the features.

## Implementation

### Data

The image data we used stems from the Caltech-256 Image Set<sup>1</sup>, which consists of 256 sets of images of a certain class. We randomly selected 5 of these, namely: cactus, dice, raccoon, spaghetti and sushi. The preparation of the image data was comprised of the following steps, which we implemented in the file `utils.py`:

- All images were resized to the same size, converted to grayscale and normalized (which removes noise from the image).
- All images are associated with a label (their image class).
- The images are split into a training set (80% of images per class) and a testing set (20% of images).

## Discussion of Results

## References

- [1] Herbert Bay, Tinnie Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, 2006.

---

<sup>1</sup>[http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/), accessed on the 22nd of December 2020