Final Project – Parking Slot Detector

Riccardo Rettore
matricola: 2110512

***Abstract*—The following project presents a possible approach to the parking slots detection problem, where the goal is to localize feasible slots inside a given image and to classify them into two different categories, the available and the occupied ones. In order to achieve this a modified-SIFT extraction algorithm and a multilayer perceptron model have been developed. Furthermore, three different regions of interest extraction techniques have been proposed, starting from a traditional approach and proceeding with a machine learning based one. Results and bounding boxes have been plotted and compared, in order to find the best approach in different frameworks and scenarios.**

Keywords—object detection, features extraction, multilayer perceptron, modified-SIFT algorithm, regions of interest proposal, computer vision

I. Introduction

To develop a parking slot detector and classifier several steps have been involved. Firstly, SIFT descriptors have been extracted and a multilayer perceptron has been trained. To keep the number of involved parameters as limited as possible, without losing information, a different SIFT approach has been presented. Multilayer perceptron model (MLP) has been trained with both traditional and modified SIFT descriptors and results have been compared. Other models have then been tested, such as support vector classifier (SVC) and logistic regression, and accuracy scores have been evaluated in order to select the best suitable model.

The second part of the project relates with the extraction of the possible regions of interest from a given image. To achieve this, three different approaches have been tested. The first one uses classical computer vision techniques to extract local patterns and possible boxes of interest. The second approach, instead, is based on machine learning techniques, and it performs an advanced selective search algorithm. Such method has been studied to compare its results with the ones from the previous one, looking for the best adaptive algorithm in different weather conditions and with different image perspectives. The third method is again a classical approach algorithm, used in a wide range of applications. It is a connected component labelling algorithm and it has been added in the case of study since it could be compared with the previous traditional approach, getting other useful results about adaptiveness of both of them. It is worth mentioning that also deep learning and convolutional network algorithms could have been used for this specific purpose, but one of the main goals of the project was to study the behaviour of traditional approaches in the proposed framework.

II. Dataset

The CNRPark+EXT dataset has been used to train and test the algorithms. It is a dataset for visual occupancy detection of parking lots of roughly 150 000 labelled images (patches) of vacant and occupied parking spaces.

The PkLot dataset could be also used. It contains 12 417 images of parking lots and 695 899 images of parking spaces segmented from them, which were manually checked and labelled. Despite its larger batch of images, it presents a more complex internal structure than CNRPark+EXT.

Hence, for the well-structured comma-separated value (.csv) file and for its vertical organizational structure, CNRPark+EXT dataset has been preferred in the MLP model training, while both CNRPark+EXT and PKLot have been used in the testing part. Furthermore, several dataset splits combinations have been made to test model's performances, studying fit time and accuracy scores obtained.

III. Training & Classification

A. SIFT descriptors extraction

The SIFT algorithm has been performed on each single patch of the split dataset. This method provide a good amount of descriptors when the slot is occupied by a car and, on the other hand, less or no descriptors when the slot is empty. The algorithm has been modified in order to standardize the number of descriptors. In this way every image has the same number of parameters, reducing the complexity of models' inputs and their training. So, instead of using the traditional 128-elements SIFT array descriptor, one extra element has been appended and an element-wise descriptors averaging has been performed. In this way each image patch has only one 129-elements array descriptor. The 129th value now contains the number of averaged descriptors. In this way a strong regularization of the number of parameters has been made but without losing too many information. The idea behind of this is that, even though SIFT descriptors represents different features, it is possible to cluster close descriptors together, obtaining a less precise description but with a more general vision. Following these reasonings, every patch has been correlated with its own descriptors.

Traditional and modified SIFT approaches have been tested with the same multilayer perceptron model (MLP). From the scores in (Table 1.) it is possible to notice how the new approach obtains slightly better values in less time with respect to the traditional SIFT algorithm.

TABLE I. COMPARISON BETWEEN TRADITIONAL AND MODIFIED SIFT ALGORITHM WITH DIFFERENT NUMBER OF IMAGES

Accuracy scores - Time needed (seconds)				
Samples	***Traditional SIFT***		***Modified SIFT***	
25 000	0.898	55 s	0.914	45 s
35 000	0.812	75 s	0.859	67 s
50 000	0.790	120 s	0.828	98 s

B. Multilayer perceptron model

The list of descriptors computed so far has been splitted into train and test samples sublists using the scikit built-in function called train_test_split. The general percentage of occupied and available slots on each subset list has been maintained. After the splitting, a multilayer perceptron model has been trained and different configurations of architectures, solvers and activation functions have been studied. From the evaluated scores (Table II.) the model with 10 neurons in one hidden layer, logistic activation_function and stochastic gradient descent (SGD) solver has been selected.

TABLE II. COMPARISON BETWEEN DIFFERENT ACTIVATION FUNCTIONS WITH 50 000 IMAGES DATASET AND SGD SOLVER

Activation functions			
	Logistic	***Relu***	***Tanh***
Best Accuracy score	0.828	0.814	0.815
Best Hidden architecture	(10,)	(30,0)	(30,10)

Since there are huge number of parameters to estimate MLP has been preferred to other techniques, as for example SVC. Indeed, as reported in Scikit docs, SVC fit time scales at least quadratically with the number of samples and may be impractical beyond tens of thousands of samples.

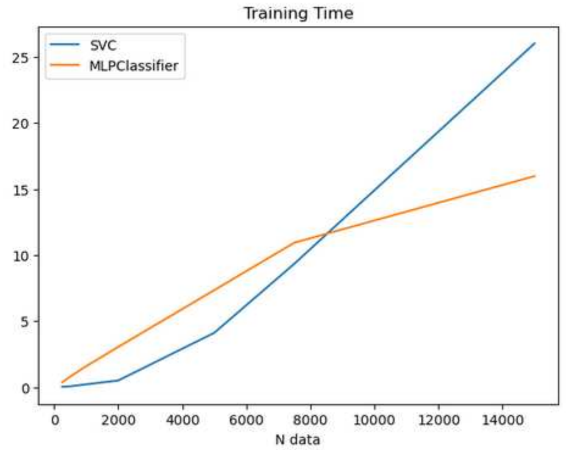


Fig. 1. Training time comparison between SVC and MLP

IV. REGIONS OF INTEREST PROPOSALS

Three different approaches have been tested. The first one uses traditional computer vision techniques to extract local patterns and then possible boxes of interest. The second approach is based on machine learning techniques and performs advanced selective search method. The third method is a connected component labelling algorithm. The resulting regions have been then classified and marked with red borders if occupied and with green borders if available.

A. *First approach*

The first approach works with classical computer vision techniques. Firstly, in order to extract relevant edges from the image, a Canny edge detector has been performed on a grey scale version of the image. This procedure has been made automatically find-tuned, since the two thresholds of Canny were estimated from an OTSU thresholding procedure. Indeed, the optimal value that comes from the latter is then used in Canny by setting the low threshold as 30% of the optimal value and the high threshold as the optimal value.

With the resulting edges a find contours function from OpenCV has been used and then a minimum area rectangle has been computed on each contour. In this way, each rectangle is described by the coordinates of the starting point, by the width, the height and by the angle rotation. In the end, the list of boxes has been filtered with several constraints, regarding area values, overlap and intersection conditions.

A final cleaning has been made by making use of a distribution percentile, keeping the boxes inside a selected range. The final list has been then used to create the patches, cropping the initial image.

Modified SIFT and MLP algorithms have been used on resulting patches, obtaining a classification for each of them. Results are then displayed in the results folder of the project.

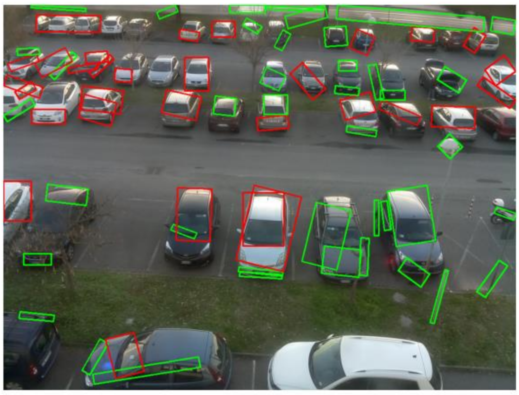


Fig. 2. Example of the results obtained from the first approach (no filtering step applied)

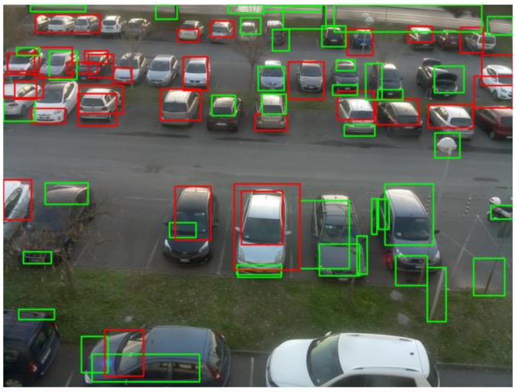


Fig. 3. Example of the same results but without minimun area algorithm (no filtering step applied)

B. *Second approach*

The selective search, a region proposal algorithm implemented in OpenCV, has been used. It uses the Felzenszwalb superpixel method to find regions of an image that could contain an object. It then merges superpixels using five keys similarity measures: color, texture, shape, size, and final-meta. For further information about how it works see [1], [2].

Fig. 4. Segmentation parameters: sigma = 0.5, K = 500, min = 50, image source: https://cs.brown.edu/people/pfelzens/segment/

The same pipeline previously described has been followed. Possible boxes have been extracted directly from the input image, sorted using incrementing area criterion and filtered with the same constraints (excessive overlapping, area bigger than the threshold). Patches have been created using the remaining boxes and modified SIFT has been computed on them. The results have been used as input of the MLP, obtaining a labelling classification for each one. Outcomes are displayed in the results folder.

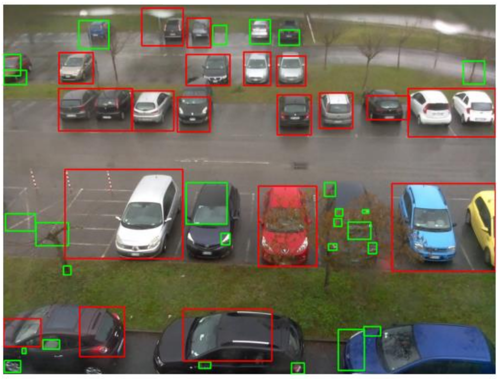


Fig. 5. Example of the results obtained from the second approach

C. Third approach

The connected component labelling algorithm, implemented in Skimage, considers two pixels as connected when they are neighbours and have the same value inside a thresholding interval. For further information see the documentation in Skimage and in the paper [3].

Input image has been pre-processed with Otsu technique and with Scikit-closing method, where dilation and erosion has been performed in order to reduce eventual salt and pepper noise and other possible noises. Regions have been extracted by applying the presented algorithm, and modified SIFT has been computed on the outcomes of previous filtering layers. Final results have been obtained using a distribution percentile. Results are then displayed in the results folder.

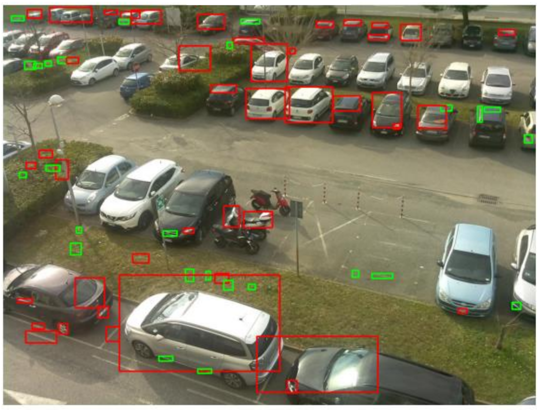


Fig. 6. Example of the results obtained from the third approach

D. Final steps

The three possible labelled boxes lists obtained so far have been merged together and then cleaned with usual overlapping constraints. Then, other optimising logics have been applied in order to reduce redundancy and to keep the number of boxes under a certain limit. Assuming as red boxes the occupied slots and as green the available ones, it is possible to state that if a red box fully contains another red one, the former is redundant and covers more space that the actual needed. So, it is possible to remove the former and maintain only the latter. It must be checked also the actual area of each box, since too small rectangles don't carry actual information. Same logic can be applied for the green boxes.

After this first cleaning it is possible to reduce even more the number of boxes by removing green boxes that overlaps too much the red ones, since in a general scenario it is preferable to mark these slots as occupied.

The initial and the optimised results are shown in the results folder.

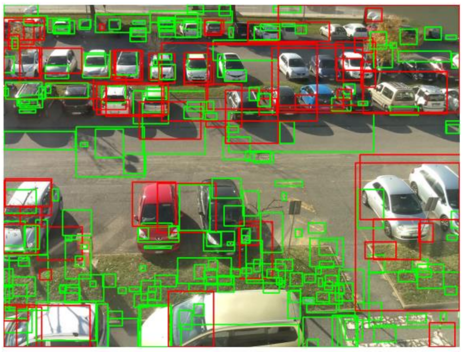


Fig. 7. Example of final merged results, before optimization

Fig. 8. Example of final merged results, after optimization

E. Other tested approaches

Customised Viola-Jones algorithm has been performed. Cascade-Trainer-Gui application has been installed and the algorithm has been trained with positive and negative folders, as written in the documentation. This approach could work with great results if correctly trained and if the images represent the full frontal view of the slots. This was not the case of study and hence the technique was not further studied. Furthermore, the datasets used to train the algorithm was not the most suitable one, since Viola-Jones algorithm requires very precise labelling in both positive and negative cases.

Histogram-based description has been tested, using its normalised results as input for a possible MLP model. Also in this case the results obtained were above the average, but some problems have been observed, mostly related with the limited resolution of the images and with the similarity between the colors of the cars and the background.

V. FINAL RESULTS AND CONCLUSIONS

The full pipeline of the project has been tested with different images and with different scenarios. Weather conditions as sunny, rainy, foggy, cloudy have been taken into consideration. Different perspectives, points of view and camera positions have been also tested.

From the plots and the results obtained it is possible to deduce several observations. The implemented classical approach works very well when the input image is not too noisy, since, in that case, contours become very unstable. To prevent this, different filtering techniques could have been used, like bilateral filters or other methods like dilation and erosion. Compared with the machine learning approach it presents better slots localization when the image perspective is not the most suitable one (frontal case). This is due to the minimum area algorithm applied, that takes into consideration also the possible rotated rectangles. With the selective search it is not possible, as writing, to have this feature and this limits its possible approaches. On the other hand, the latter presents very good results in the selection of possible meaningful rectangles. The traditional based approach, instead, presents some limitations in this case, especially due to how it has been developed.

Regarding the connected component algorithm, it is possible to notice how the results are very limited, since it is not very suitable to work with this environment.

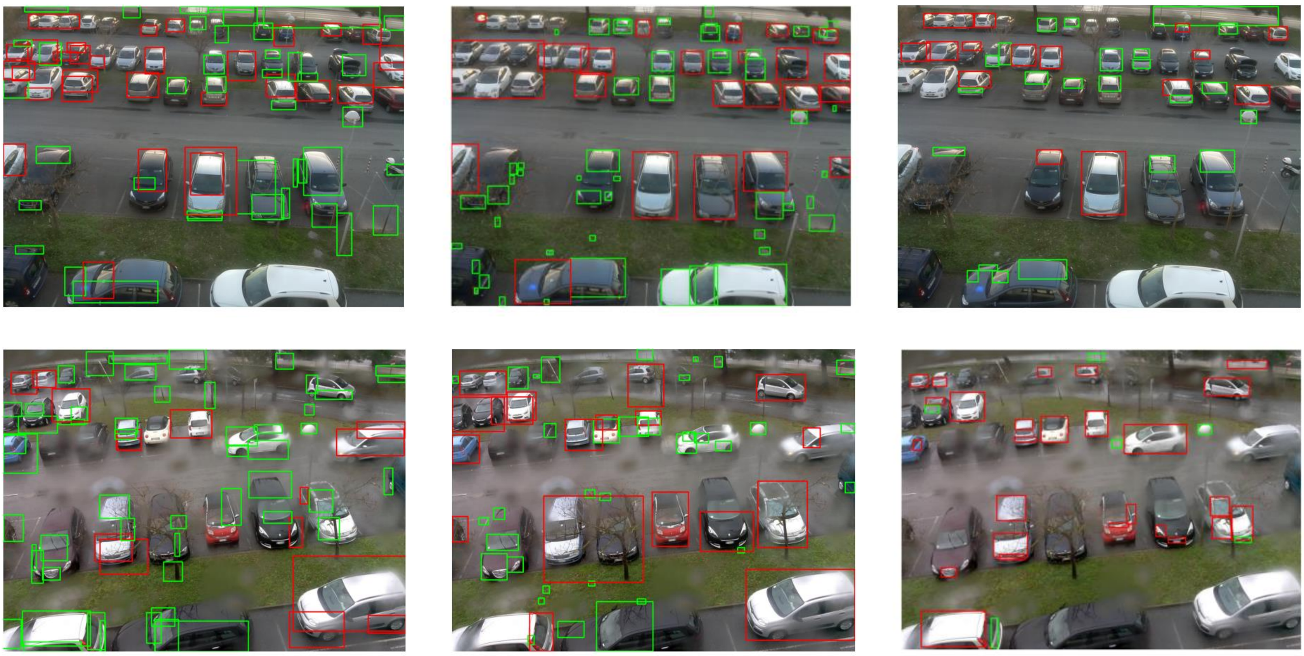
Comparing the different weather conditions, it is possible to notice that the first two approaches maintain stable results in each scenario and they complete each other in most of the cases. A good choice could be that of merging them together and to keep the overall optimised result.

Considerations about the modified SIFT algorithm can be made. Results obtained from testing traditional SIFT and modified SIFT highlights better accuracy scores with less fit time in the second case. However, there is no way to prove it in a general framework. To conclude, other training techniques have been tested, like SVC and logistic regression, but the scores proved to be unsatisfactory with respect to the MLP model.

It is worth mentioning that with the implemented algorithm for the regions extraction there is a space flattening problem behind. Indeed, boxes bring no space meaning of what they have inside, but they just represent a possible region of interest. So, when the background, or everything near the parking slots, has a relevant part in the full image it could happen that the algorithm marks as possible regions even there. Hence, further upgrades of the code should take this problem into account by using suitable labelling features descriptor to mark each single box before the classification phase.

To conclude, PKlot images have been tested and the obtained results confirmed a good generalization behaviour of the algorithm. Further studies could swap the datasets, training the MLP with Pklot and testing the model with CNRPark+EXT.

Fig. 9. Example of final merged results, after optimization, on PKLot dataset image



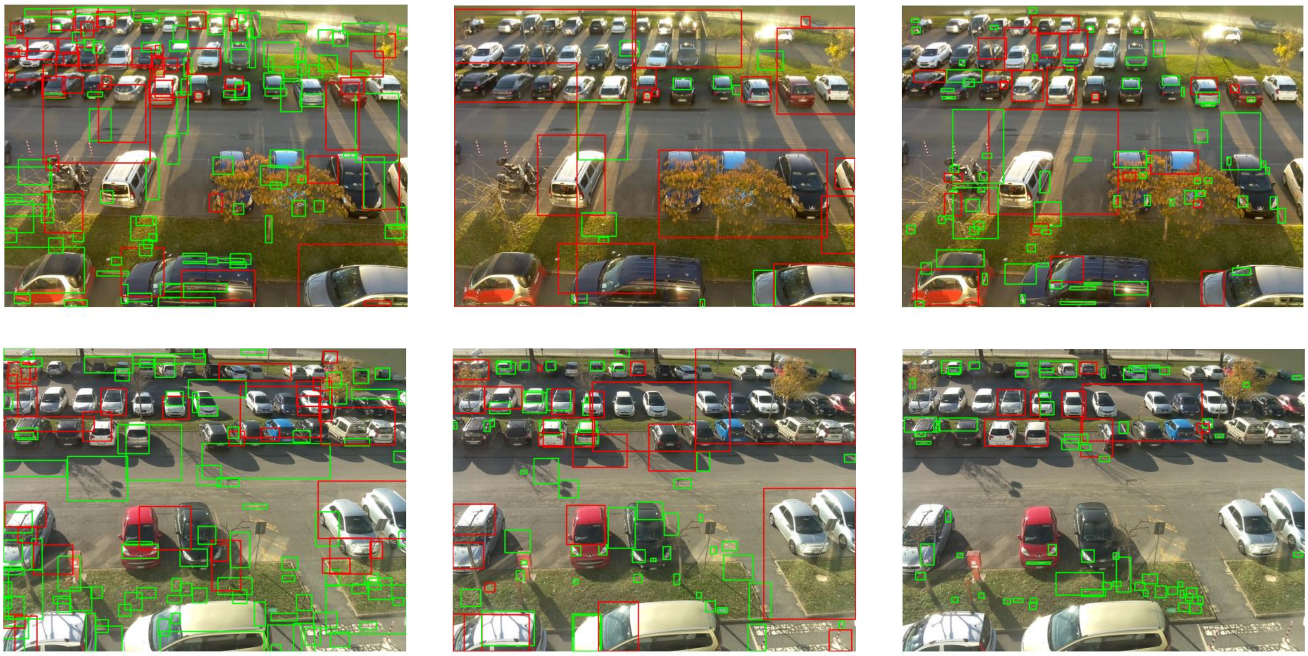


Fig. 10. Comparison between the three approaches with different weather conditions. From left to right: developed approach, machine learning based approach, connected component approach

REFERENCES

[1] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders, “Selective Search for Object Recognition”, Technical Report 2012, submitted to IJCV

[2] Pedro F. Felzenszwalb, Daniel P. Huttenlocher, “Efficient Graph-Based Image Segmentation” paper.

[3] Kesheng Wu, Ekow Otoo and Arie Shoshani Lawrence Berkeley National Laboratory, Berkeley, CA, USA , “Optimizing connected component labeling algorithms” paper.