FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
MIEIC - 2013/2014

COMPUTER VISION

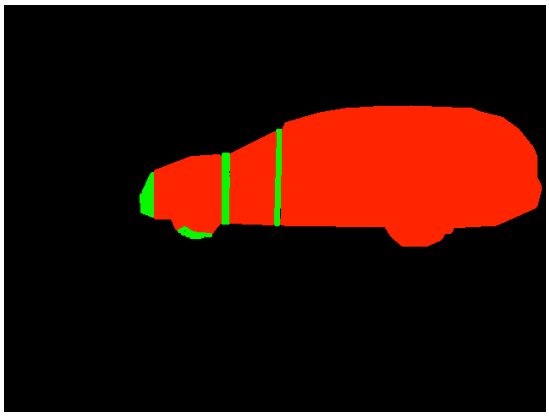
PROJECT Nº 2
Car detection in images

Summary

The main goal of this project is to develop a system that can detect cars in images, giving an estimate of the cars' location in the image. The system should be based in a classifier trained to distinguish cars from any other visual content. The description scheme should be based on the *bag-of-words* approach.

Dataset

The images comprising the dataset for this work are available in Moodle. The dataset is divided into training and testing images. It is based in the Graz-02 dataset with INRIA annotations.



An example from the Graz-02 images of cars is shown above. The original images are accompanied with ground-truth masks indicating the image regions occupied by cars. Red colour marks the visible object parts and green is used for occlusions. Each image can have more than one mask associated with it.

Recognition system

The recognition system should include:

- **Feature detection**: A local interest point detector is used to find relevant features.
- **Feature description**: A local descriptor scheme is used to extract an invariant descriptor.
- **Visual vocabulary/dictionary**: All the descriptors extracted from all the images are clustered (e.g. k-means) to quantize the space into "visual words". In other words, each cluster mean represents a word of the resulting dictionary.
- ***Bag-of-words representation***: A given area is represented by a histogram of words, using the vocabulary built previously. The resulting bag-of-words representation vector should be normalized.
- **Training**: A binary classifier should be trained using only the training images of the dataset. Consider that areas in the visible object mask (coloured in red) are part of the "car" class and that all other areas are part of the "non-car" class. The model should be trained with a balanced number of examples of both classes.
- **Testing**: Since the location and scale of the cars are not known a priori it is necessary to analyse a range of image locations and scales. A scanning window procedure should be used in this case. Each window should be represented by a bag-of-words vector and tested with the previously trained model. The result will be the areas of the image containing a car, possibly defined by bounding boxes.

Extra:

- **Evaluation**: Compare at least two different approaches, for example, different classifiers, different local descriptors or combining other descriptors studied in the Computer Vision course. To evaluate the performance in an image, compare the location of the car given by your system with the mask provided in the dataset. If more than a given percentage of the pixels of the mask are inside the bounding box, consider it a correct identification (true positive); otherwise, consider it an incorrect identification (false positive). Only the testing images should be used in the evaluation.
- Other improvements will be considered in the evaluation, if justified.

Scientific Paper and Delivery

A short report should be elaborated in the format of a scientific paper (max. 4 pages), including:

- Brief introduction to the problem, including references about the state of the art;
- Description of the developed system;
- Possible additional specifications or improvements;
- Results of the object recognition, namely percentage of objects correctly identified and other measures considered relevant (e.g: true positive rate, true negative rate, accuracy, etc.);
- Discussion about the overall performance of the system and possible situations where it fails;
- (extra) Comparison of performance using different approaches;
- Conclusions and future improvements.

The paper can be written in English or Portuguese and should be based on the model available in Moodle. The code, with meaningful comments, should be presented in annex.

The work should be submitted at the Computer Vision page, in the UP Moodle site, until the end of the day December 16, 2013.

Bibliography and other support material

- Visual Categorization with Bags of Keypoints, C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, ECCV International Workshop on Statistical Learning in Computer Vision, 2004.
- Sampling Strategies for Bag-of-Features Image Classification. E. Nowak, F. Jurie, and B. Triggs. ECCV 2006.
- Video Google: A Text Retrieval Approach to Object Matching in Videos, J. Sivic and A. Zisserman, ICCV 2003.
- Object Recognition from Local Scale-Invariant Features, D. Lowe, ICCV 1999.