

PROJECT Nº 2

Object Recognition

Summary

The main goal of this project is to develop a system that can classify an image as being one of 10 different classes. This project is based in a past Kaggle competition available at <https://www.kaggle.com/c/cifar-10>

Object recognition system

There is no restriction on how the system should be built. Nevertheless, a typical object recognition system based in the Bag of Words (BoW) approach, includes the following modules:

- **Feature extraction:** A local interest point detector and descriptor is used to find relevant features and extract the corresponding descriptors.
- **Image representation:** All the descriptors extracted from all the images are clustered (e.g. k-means) to quantize the space into “visual words”. Each cluster mean represents a word of the resulting dictionary. An image is then represented by a histogram of words, using the vocabulary built previously. The resulting bag-of-words representation vector should be normalized.
- **Classification:** A classifier is trained with the training images defined for the dataset. After having the model trained, the input image is tested in order to obtain a category. Since the number of images is very large, a reduced dataset may be used, if necessary.

Additional modules can be implemented and will be evaluated as extras:

- **Evaluation:** Compare at least two different approaches, for example, different classifiers or different local descriptors studied in the Computer Vision course. Only a set of testing images should be used in the evaluation.
- **Deep learning:** Try to use a deep learning methodology to implement the system. Note that the Keras library (<http://keras.io>) already includes the CIFAR-10 dataset. Take also into consideration that the training time will take a significant amount of time if GPUs are not used.
- Other improvements will be considered in the evaluation, if justified.

Scientific Paper and Delivery

A short report must be elaborated in the format of a scientific paper (max. 3 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 image retrieval system, namely percentage of categories correctly identified and other measures considered relevant (e.g: confusion matrix, 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 must be submitted at the Computer Vision page, in the UP Moodle site, until the end of the day December 12, 2016.

Bibliography and other support material

- Video Google: A Text Retrieval Approach to Object Matching in Videos, J. Sivic and A. Zisserman, ICCV 2003.
- Sampling Strategies for Bag-of-Features Image Classification. E. Nowak, F. Jurie, and B. Triggs. ECCV 2006.
- Imagenet classification with deep convolutional neural networks, A. Krizhevsky, I. Sutskever, and G. E. Hinton. *Advances in neural information processing systems*, pp. 1097-1105. 2012.