http://www.cs.ucf.edu/~bagci

[Programming Assignment] (3)

Robot Vision

Dr. Ulas Bagci • (Spring) 2019 • University of Central Florida (UCF)

Coding Standard and General Requirements

Code for all programming assignments should be **well documented**. A working program with no comments will receive **only partial credit**. Documentation entails writing a description of each function/method, class/structure, as well as comments throughout the code to explain the program flow. Programming language for the assignment is **Python**.

Following libraries can be used extensively throughout the course:

- PIL (The Python Imaging Library), Matplotlib, NumPy, SciPy, LibSVM, OpenCV, VLFeat, python-graph, TensorFlow, PyTorch.

If you use CANOPY, make sure that you use version 2.7 (or whatever your TA suggests), which already includes many libraries. If you are asked to implement "Gaussian Filtering", you are not allowed to use a Gaussian function from a known library, you need to implement it from scratch.

Submit by **25th of April 2019**, 11.59pm.

Action Recognition [15 pts]

You task is to recognize actions from videos using machine learning classifier(s) and suitable vision features. You will use UCF sports action data set here`http://crcv.ucf.edu/data/ucf_sports_actions.zip`. UCF Sports dataset consists of a set of actions collected from various sports which are typically featured on broadcast television channels such as the BBC and ESPN. The video sequences were obtained from a wide range of stock footage websites including BBC Motion gallery and GettyImages. The dataset includes a total of 150 sequences with the resolution of 720 x 480. The collection represents a natural pool of actions featured in a wide range of scenes and viewpoints. By releasing the data set we hope to encourage further research into this class of action recognition in unconstrained environments. Since its introduction, the dataset has been used for numerous applications such as: action recognition, action localization, and saliency detection.

The dataset includes the following 10 actions.

Diving (14 videos)
Golf Swing (18 videos)
Kicking (20 videos)
Lifting (6 videos)
Riding Horse (12 videos)
Running (13 videos)
SkateBoarding (12 videos)
Swing-Bench (20 videos)

Swing-Side (13 videos)
Walking (22 videos)

Feature Extraction [3 pts]: You have full freedom to design feature extraction technique, optimize it with respect to your settings, and evaluate its success. You are required to report parameters and details of your feature extraction method. Also, mention the reason behind the choice of features that you are using. If you are not using deep learning methodologies, you are suggested to use HOG (Histogram of Gradient) features at least.
Note: To work with UCF sports dataset, you need to write your own data-loader to read video frames and action labels. See `https://pytorch.org/tutorials/beginner/data_loading_tutorial.html` for sample data-loader.

Classifier Design [7 pts]: Support Vector Machines (SVM), Random Decision Forests (RF), or Neural Networks (NN) (including CNN) can be used/designed to solve the task being asked. Try to understand how classifier works, train it properly before doing test. In case you choose NNs, You may choose to train an end-to-end frame-level action classifier (e.g. VGG, ResNet, etc), in which case the accuracy should be reported as the average or majority voting on all video frames for each test video.
You are required to report parameters of your classifier in the evaluation section.

Evaluation [5 pts]: For Action Recognition, Leave-One-Out (LOO) cross-validation scheme is recommended. This scenario takes out one sample video for testing and trains using all of the remaining videos of an action class. This is performed for every sample video in a cyclic manner, and the overall accuracy is obtained by averaging the accuracy of all iterations. Note that all video frames have ground truth labelled. Sensitivity, specificity, and accuracy should be included in the evaluation.You are expected to write one page summary, describing all steps (feature extraction, classifier design, and evaluation).

Reference paper: Mikel D. Rodriguez, Javed Ahmed, and Mubarak Shah, Action MACH: A Spatio-temporal Maximum Average Correlation Height Filter for Action Recognition, Computer Vision and Pattern Recognition, 2008.