The project attached to this report refers to a hand detection and segmentation task, a very important task which is useful in varius modern application, in particular for applications of human-robot interaction, gesture recognition or human activity analysis.

The overall project can be seen as a combination of two separeta subprojects: hand detection and hand segmentation.

# Hand Detection

For the first step we tried to use the concept of Bag of Words, which classifies an image by using a vocabulary of “words” that in this case correspond s to features. In generale, in document classification a bag of words is a sparse vector of occurrence counts of words, that is a sparse histogram over vocabulary, while in computer vision, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features. This strategy relies mainly on three steps;

1. Feature detection, which selects and highlights the most meaningful points in the image. In particular the selected image is of a detailed object that we want to detect in the final steps;
2. Feature description, which are vectors describing the area surrounding each keypoint detected in the previous step. Tipically this two steps can be obtained with SIFT (Scale Invariant Feature Transform), which converts each patch to a 128-dimensional vector;
3. Codebook generation, the core of the algorithm whichconverts the vector represented patches to “codewords”, considerate as a representative of several similar patches. To implement this part is need a clustering technique, like kmeans clustering, in order to group similar patches to a one large single patch. Codewords are defined as the center of the learned clusters and the image can be represented by the histogram of codewords;

However, we noticed that even if the algorithm has good time and computational complexity performance, the overall results didn’t match with the requirements, because we only managed to classify images as with hand or without hand and no findind the location on the image, so we keeped this part only for the first evaluation of the project in order to quickly discard images without hands.

For the core of this first part we decided to work with a boosted cascade of weak classifiers, which is divided in two stages: the training and the detection stage.

For training a boosted cascade of weak classifiers we need a set of positive samples, containing the hands, and a set of negative images, which corresponds tipically to the background of the images used for the positive samples.