Random Forests are one of the strongest modern methods of machine learning and are applied in many different applications. The basic idea is to generate many decision trees, where tree creation is subject to a certain amount of randomization. The final system answer is the fusion of the result of the individual trees. In principle, Random Forests can be interpreted as a special kind of nearest-neighbor-search where the distance measure is generic and task-dependend optimized instead of predefined.

Each data point follows a specific path through the tree. The more similar two samples are, the more likely will their paths overlap. The overlap of their paths can therefore be used as a similarity measure, which is automatically learnt from the provided data and optimized to the task at hand.

The goal of this thesis is to implement a Random Forest, which is taylored to the application to compute distances/similarities between data points. The derived distance measure shall be used in an example computer vision task such as clustering, image retrieval, classification, or keypoint matching.

Keywords: Random Forests, clustering, kernel methods

Involved tasks:
– Literature Research
– Extension of existing frameworks or reimplementation of Random Forests
– Evaluation of the proposed approach based on an example computer vision task

(Recommended) requirements:
– Good knowledge about digital image processing (e.g. attendance in lecture DIP)
– Good knowledge about machine learning (e.g. attendance in lecture AIA)
– Good programming skills (e.g. C++)

Language: German / English