THE MASH PROJECT

François Fleuret[(1)], Philip Abbet[(1)], Charles Dubout[(1)], Leonidas Lefakis[(1)], Alexandre Nanchen[(1)], Gilles Blanchard[(2)], Nicole Krämer[(2)], Andre Beinrucker[(2)], Olivier Teytaud[(3)], Nataliya Sokolovska[(3)], Jean-Baptiste Hoock[(3)], Yves Grandvalet[(4)], Gérard Govaert[(4)], Xiao Liu[(4)], Luis Francisco Sánchez[(4)], Jiri Matas[(5)], Vladimir Smutny[(5)], and Martin Dubec[(5)]

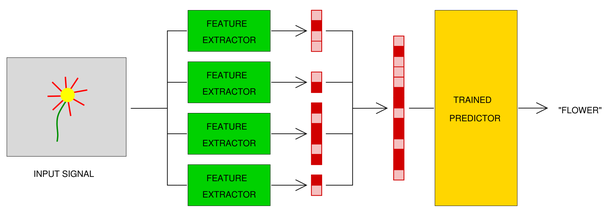
(1) Idiap Research Institute, (2) Weierstrass Institute for Applied Analysis and Stochastics, (3) INRIA, (4) CNRS, (5) Czech Technical University in Prague

INTRODUCTION

The MASH project is an attempt at coordinating a large team around the design of an extended learning system. It aims at developing new tools for the collaborative design of a very large family of feature extractors, and at studying the performance of the resulting system on object detection and goal planning.

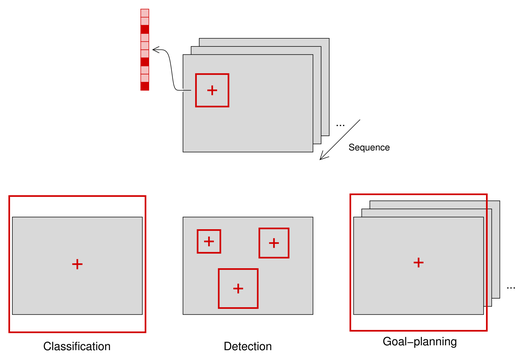
FEATURE EXTRACTION

The project focuses on standard combinations of feature extractors and ML methods.



HEURISTICS

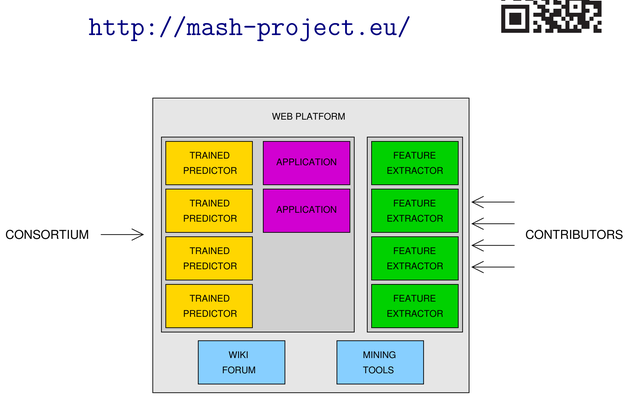
We define the concept of a *heuristic*, a feature extractor with a persistent state.



OPEN PLATFORM

The project activities are organized on a web platform accessible at

`http://mash-project.eu/`



The consortium provides and controls the prediction methods, databases, goal-planning tasks, and analysis tools.

Any contributor can upload C++ implementations of heuristics:

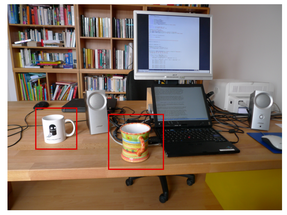
```
void init();
unsigned int dim();
void prepareForImage();
void finishForImage();
void prepareForCoordinates();
void finishForCoordinates();
scalar_t computeFeature(unsigned int index);
```

The platform runs experiments and distributes computation over multiple computers on different geographic sites.

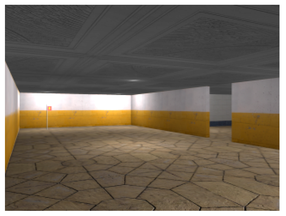
Heuristics are run in a sand-boxed environment, with controlled memory and CPU usage.

APPLICATIONS

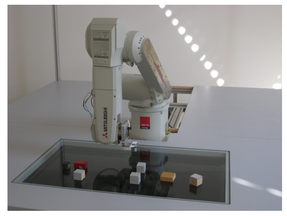
Image classification



Object detection



3D simulation



Robotic arm

ON-LINE TOOLS

The platform hosts:

- Rich documentation, including video tutorials.
- Source code repository.
- Private and public heuristics.
- Tracking of heuristic phylogeny.
- Wiki, forums, and messaging.
- Contributor-tailored experiments.
- Public large-scale experiments.
- Summaries of the worst mistakes.

The machine learning core includes multiple prediction schemes to combine the heuristics: AdaBoost, C4.5, Linear SVM, Perceptron, Kernel Perceptron.

CHALLENGES

Coordinating and motivating multiple contributors involves multiple challenges:

- Over-fitting: The feature space is huge.
- *Meta* over-fitting: Contributors will design features specific to the problems at hand.
- Fairness: Contributions must be judged on an equal foot.
- Computational cost: (Pre-)processing cannot be factorized as much as usually.
- Security: We have to compile and execute alien code on the computers of the project.
- Intellectual property: The resulting system integrates code from multiple origins.

RESEARCH

The project investigates three main axes:

- Classification and regression with a large and heterogeneous feature space.
- Goal-planning with a large and heterogeneous feature space.
- Clustering of heuristics and mistakes.

CONSORTIUM

- IDIAP (Switzerland)
- WIAS (Germany)
- INRIA (France)
- CNRS (France)
- CVUT (Czech Republic)

Contact: `francois.fleuret@idiap.ch`