Image processing and scientific workflows in the Virtual Observatory context

Eric Slezak[1], André Schaaff[2] and Jean-Julien Claudon[2]

[1]Observatoire de la Côte d'Azur, BP 4229, F-06304 Nice cedex 4, France
email: slezak@obs-nice.fr

[2]Observatoire de Strasbourg, 11 rue de l'Université, F-67000 Strasbourg, France
email: schaaff@astro.u-strasbg.fr

Since 2005, the workflow working group of OV-France gathers astronomers from five french institutes with common objectives: (*i*) defining use cases of general interest; (*ii*) identifying the simplest workflow structure allowing portability; and (*iii*) suggesting solutions for designing and exploiting easily such workflows (cf. Astrogrid developments). Our workflow definition is the following: a sequence of tasks executed within a controlled context by an architecture taking into account VO standards.

The analysis of multivariate images is one of these uses cases requiring advanced algorithms. These data are indeed intrinsically complex due to potential non-gaussian behaviours, to their inherent sparsity and to hidden correlations. Basically, three main reduction tasks have to be carefully performed to properly handle them. First, one has to reduce the dimensionality of the data space by selecting or building variables carrying the relevant information. Then, after the choice of a metric and a model allowing a sparse representation, data are segmented and, finally, primitives are extracted and classification achieved. Let us consider for instance the multisource data fusion problem. The goal is to combine in an optimal way all the data into a single model while preserving all the information. To do so, different strategies are possible considering either a set of 2D images, a set of spectra or a data cube, as well as different tactics for each of them like for instance a chi-square sum or a bayesian approach for co-adding the images. Therefore we are developing new algorithms (MDA project) relevant to such data sets as well as related VO compliant web services.

To achieve a particular goal, several of these tools must in addition be organised in a rigorous plan including detailed knowledge about the use of each program and about the overall task. Such workflows encode a know-how information for solving complex tasks which has also to be made available to the end-user. For this purpose, we are developing the AIDA framework for graphically building and executing analysis systems from elementary processing blocks. Since the AIDA architecture itself is described in the next paper of this special session, let us just summarize two of its features from the user's point-of-view. First, the integration of each (local or distant) program in the AIDA repository is straightforward since it only requires to upload a file describing the number and type of the I/Os and another one storing the values of the variables needed by the algorithm; each code is then automatically wrapped to become accessible through CGI and Web Services. Second, to build any workflow, one has only to chain programs by (i) dragging them from the repository to the board using the mouse for selecting them and (ii) interconnecting the inputs/outputs at the edges of their graphical icons with a line (compatibility checks are automatically performed). A genuine action plan is built in this (graphical) way, which can be executed on the local computer running the AIDA workflow engine.