

Intelligent Video Analysis for Image Processing-Naive Users

Gaya Nadarajan, Jessica Chen-Burger, Robert B. Fisher

gaya.n@ed.ac.uk

http://homepages.inf.ed.ac.uk/s0450937

University of Edinburgh, U.K.



Abstract: Video data collected continuously are pervasive today but analysing them in an efficient manner has proven to be a challenge: 1) Manual processing is too time-consuming. 2) Highly specialised computational tools are often not available. Automated assistance would speed this up considerably, however, would require the user to write complex image processing (IP) programs. To overcome this, a combination of planning and ontologies are used to select a set of performance-based IP tools within a workflow composition and execution framework. This approach capitalises on user-provided goals, constraints and domain knowledge as well as IP experts' heuristics (best-practices) to provide a *fast* mechanism of generating solutions. Experiments were conducted for video classification, fish detection and counting tasks in unconstrained underwater videos.

Key Requirements

A set of requirements are formulated to address the problem of automatic video analysis for IP-naive users:

- Process Automation
- User Requirements and Domain Description Capture
- Rich Process Modelling for VIP Tasks
- Performance-Based IP Tool Selection
- Adaptive, Flexible and Generic Architecture

Existing efforts in workflow, image processing and knowledge engineering communities, however, fail to address all these requirements.

The Approach

A hybrid **workflow composition and execution** mechanism within a three-layered framework is implemented.

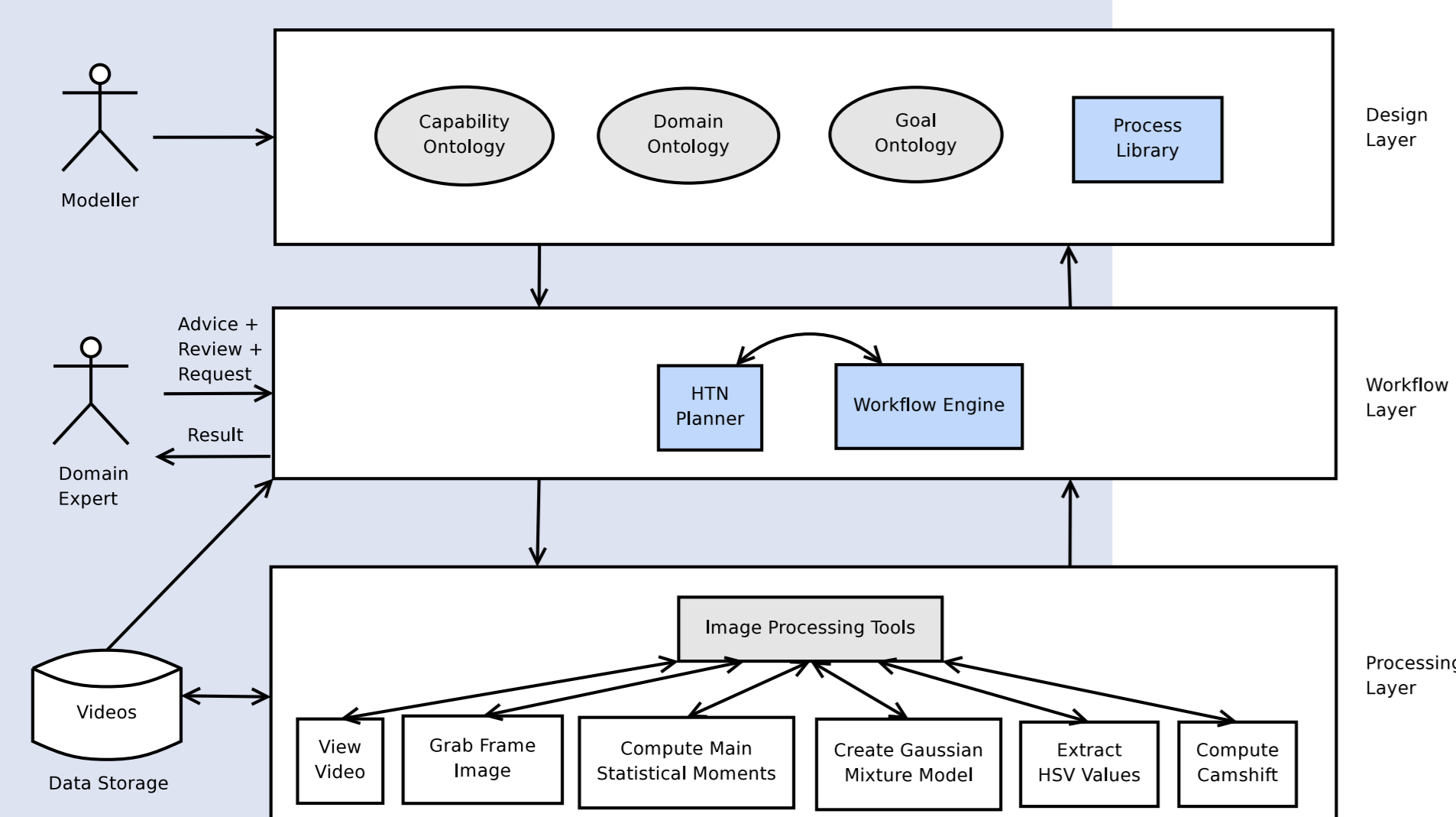


Fig. 1 Hybrid workflow composition and execution framework for automatic video analysis for IP-naive users

Ontologies

The video and image processing (VIP) ontology is modularised into three sub-ontologies to describe the concepts and relationships in the domain, as well as for inference. The ontologies were constructed and used for goal and initial domain description (to assist the workflow) formulation and performance-tool selection (to assist the planner).

- **Goal** ontology contains the high-level VIP tasks (goals) and constraints that the user will communicate to the system (example: see Fig. 2).
- **Video and Image** ontology describes the concepts and relationships associated with the videos, such as the lighting conditions, colour information, position, orientation as well as spatial and temporal aspects.
- **Capability** ontology contains the classes of VIP tools, the functions that they can accomplish and the conditions for these (e.g. tied to the domain description and/or constraints).

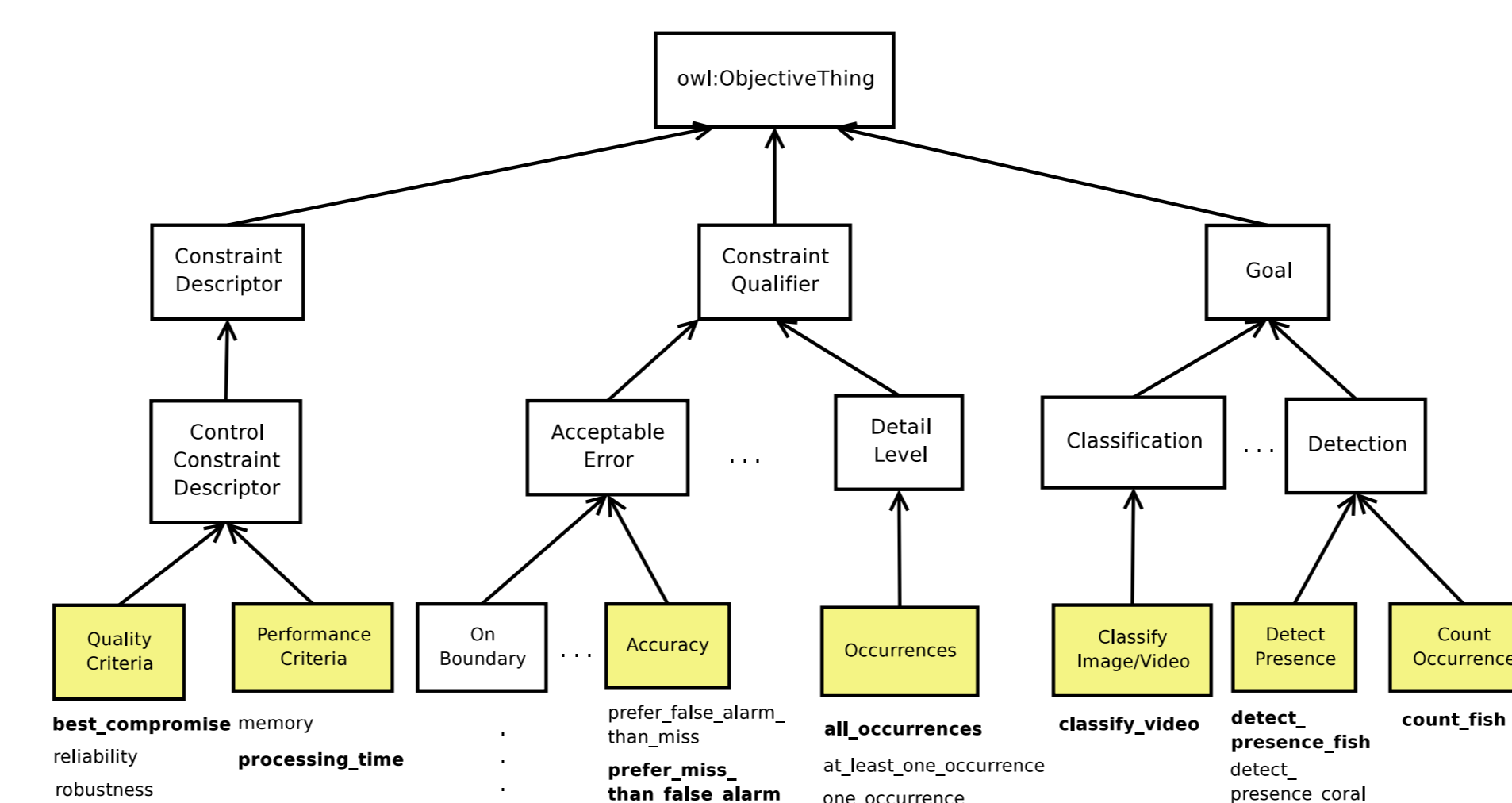


Fig. 2 Goal ontology with relevant classes and instances for the task "classify video, detect fish and count fish".

Planning

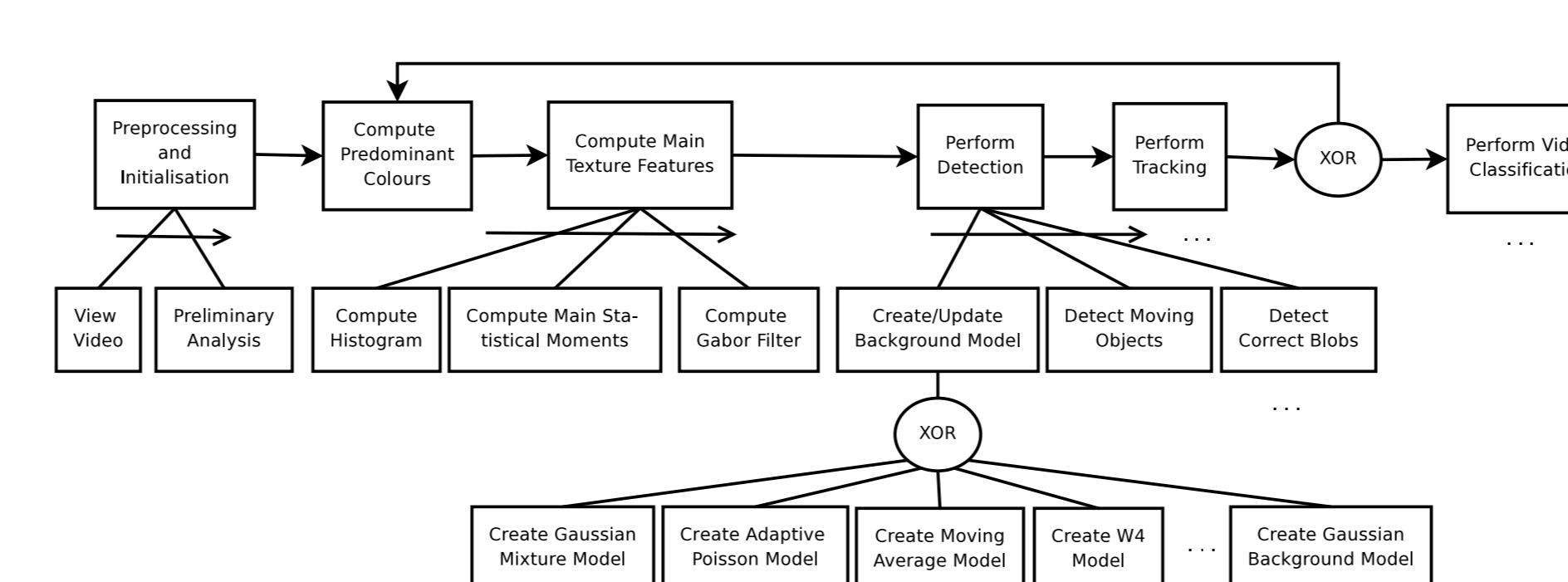


Fig. 3 Top level plan for "classify video, detect fish and count fish".

A VIP solution generation is seen as a planning problem where the goal is to perform a set of tasks (Fig. 3). Adopting HTN planning, tasks are *decomposed into subtasks exhaustively until primitive tasks are encountered*. The final outcome is a set of operators (IP tools) that can

perform the primitive tasks and their parameters. This provides an efficient way of searching for solutions by only considering the applicable methods (matching preconditions) to the current goal. In cases where more than one tool is available to perform a task, the user makes a selection, assisted by the inference from the capability ontology on the tools.

Integration with Workflow

The workflow prompts the user for video data, goal, constraints and video description. Default constraints are assigned and video description are computed automatically in the absence of user input. Assisted by the goal and video and image ontologies, the input to the planner is formulated. Each step of the way, the planner derives an IP operator for the primitive task using preconditions and inference from the capability ontology, executes it and applies the effects to the current state. The methods, primitive tasks, their preconditions and effects are encoded in a process library. User is given the resulting text and visual output of the steps and the final result, normally a video file annotated with the results (see Fig. 4). Based on that they may choose to 1) Execute the derived solution again. 2) Modify the constraints and/or video description for the same video and task (replan). 3) Evaluate the performance of the derived solution. 4) Return the main menu and start over.



Fig. 4 Sample annotated video captures for video classification, fish detection and fish counting task.

Conclusions

A hybrid workflow approach that utilises planning and ontologies is useful to construct VIP solutions efficiently for IP-naive users. Mechanisms to learn the HTN methods would ease the burden of domain modelling in the future.