

# A Modular Approach for Automating Video Analysis

Gayathri Nadarajan<sup>1</sup> and Arnaud Renouf<sup>2</sup>

<sup>1</sup> CISA, School of Informatics, University of Edinburgh, Scotland  
gaya.n@ed.ac.uk

<sup>2</sup> GREYC Laboratory (CNRS UMR 6072), Caen cedex France  
Arnaud.Renouf@greyc.ensicaen.fr

**Abstract.** Automating the steps involved in video processing has yet to be tackled with much success by vision developers and knowledge engineers. This is due to the difficulty in formulating vision problems and their solutions in a generalised manner. In this collaborated work, we introduce a modular approach that utilises ontologies to capture the goals, domain description and capabilities for performing video analysis. This modularisation is tested on real-world videos from an ecological source and proves useful in conceptualising and generalising video processing tasks. On a more significant note, this could be used in a framework for automatic video analysis in emerging infrastructures such as the Grid.

**Key words:** Knowledge-Based Vision, Ontological Engineering, Automatic Video Analysis, Ontology-Based Systems

## 1 Introduction

The field of video analysis is becoming more and more important with the fast advancement in vision technologies and the increasing size of real-time data that need to be processed efficiently. Although the majority of vision developers focus on improving low-level techniques and algorithms that perform with extreme accuracy, there is a lack of effort in conceptualising the tasks involved in the process of video analysis itself, although much of image processing involves sequences of repeated sub-tasks. Ontological engineering [1], on the other hand, is concerned with providing formal conceptualisations of entities that are relevant to a particular problem domain so that these representations and the relationships between them are made explicit. Applying higher level knowledge or semantics would allow for better reasoning on the concepts by the sharing and reuse of existing knowledge and also lead to the discovery of new knowledge.

To date, the problem of conceptualising image processing solutions has been tackled by the vision community in limited ways [2–6] and less addressed by the knowledge engineering community. However, this poses difficulty to vision-based researchers as they lack the expertise to modularise image processing problems using semantic-based approaches (e.g. ontologies). Thus a collaborated effort between ontology engineers and vision developers could bridge this current research

gap that both fields have yet to tackle with much success. The main challenges lie in the fact that the most recent knowledge representation and reasoning methods (e.g. Semantic Web technologies, such as OWL [7]) lack maturity and the image processing formulation is hard to generalise [8, 9].

In this work we tackle the problem of providing a modularised approach for video analysis using an example test case from the EcoGrid project [10]. This effort improves and extends existing work by the authors [9, 11] by adding process modelling aspects, richer modularisation of ontologies while making use of extensive video processing terminologies. The long-term effect of this is to provide a semantic-based automatic video analysis framework that is flexible which would prove useful for emerging infrastructures such as the Grid [12].

## 2 Related Work

Attempts to solve automatically image processing problems were conducted within knowledge-based systems such as LLVE [8], CONNY [13], OCAPI [14], MVP [15] and BORG [16]. However these systems remain limited to a list of restricted and well known goals. Therefore a priori knowledge on the application context (domain-specific concepts such as sensor type, noise, lighting, etc) and on the goal to achieve were implicitly encoded in the knowledge base. This implicit knowledge restricts the range of application domains for these systems and it is one of the reasons for their failure [17].

More recent approaches bring more explicit modelling [2–6] but they are all restricted to the modelling of business objects description for tasks such as detection, segmentation, image retrieval, image annotation or recognition applications. They use ontologies that provide the concepts needed for this description (a visual concept ontology for object recognition in [3, 5], a visual descriptor ontology for semantic annotation of images and videos in [18] or image processing primitives in [2]) or they capture the business knowledge through meetings with the specialists (e.g. use of the NIAM/ORM method in [4] to collect and map the business knowledge to the vision knowledge). But they do not completely tackle the problem of the application context description (just briefly in [3, 5]) and the effects of this context on the images (environment, lighting, sensor, image format). Moreover they do not define the means to describe the image content when objects are a priori unknown or unusable, for instance in robotics, image retrieval or restoration applications. They also suppose that the objectives are well known (to detect, to extract or to recognise an object with a restricted set of constraints) and therefore they do not address their specification.

To overcome these limitations, we aim at designing a modular approach for automating video and image processing. In such an approach, we have to make the formulation of the problem to be solved and the knowledge used by image processing experts during the design of the solution *explicit*. Vision experts design applications through trial-and-error cycles by implementing new solutions from scratch each time instead of reusing already developed ones. The lack of application formulation and modelling is a reason for this. Image processing ex-

perts do not realise a complete and rigorous formulation of the applications. Therefore the reusability of the applications is very poor and modularisation is needed to improve this situation.

### 3 Modularisation

Our modularisation of the image processing field clearly separates the formulation of the problems and the solutions to be produced. The user, represented by domain experts, will be involved in providing the problem descriptions in high level terms while the system (utilising a Planner) will provide the solutions in low level vision terms. Domain knowledge has to be acquired from the domain expert and formalised in order to be able to propose a way of processing the images taken under consideration.

Our first work [9] studied the formulation of image processing applications in order to propose an ontology that is used to express the objective of the domain expert (the goal part of the ontology) and define the image class to be processed (the input images and their variability description using the domain part of the ontology). It is used in the Hermes Project [19] which proposes a human-machine interface dedicated to domain experts inexperienced in the image processing field. Using this interface, they are able to formulate their goals and the description of their images using their domain knowledge.

The goal, coupled with the domain description, will then be used to derive a set of image processing tasks that will solve the user request. Each task could be further decomposed into sub-tasks. We associate an image processing tool as possessing the capability to achieve a sub-task.

In our current approach, we have opted to incorporate three ontologies to separate the goals from the capabilities and to provide meaning for the process within a semantically integrated system. Each ontology holds a vocabulary of classes of things that it represents and the relationships between them. The use of ontologies is beneficial because they provide a formal and explicit means to represent concepts, relationships and properties in a domain. They play an important role in fulfilling semantic interoperability, as highlighted in Section 1. A system with full ontological integration has several advantages. It allows for cross-checking between ontologies, addition of new concepts into the system and discovery of new knowledge within the system. A framework incorporating the ontologies with a Planning mechanism is proposed in [11].

#### 3.1 Goal Ontology

The goal ontology contains the high level goals and constraints that the user will communicate to the system. These are represented by the concepts **Goal**, **Constraint Category**, **Constraint Descriptor** and **Constraint Qualifier** in Fig. 1. The concept **Task/Process** links the goal to the processes that are associated with it. The instances of processes are contained within a process library [11] and will be selected based on task decomposition and performance

criteria. These tasks will also be linked with the capability ontology (Section 3.3).

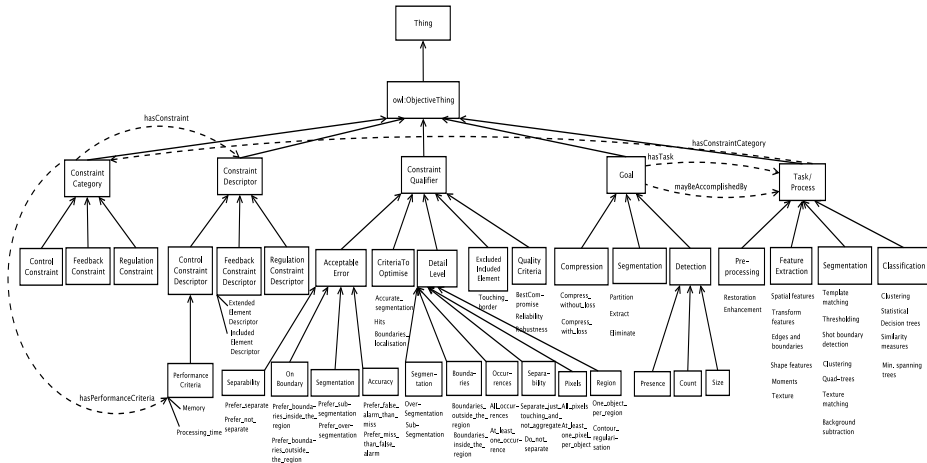


Fig. 1. Goal Ontology

### 3.2 Domain Ontology

The domain ontology (Fig. 2) describes the concepts and relationships of the application area, such as the lighting conditions, colour information, position, orientation as well as spatial and temporal aspects. The user will input the domain description along with the goal of the problem. The system will use this ontology to build the user request based on the domain description before feeding it into the Planner which will be responsible for the solution generation.

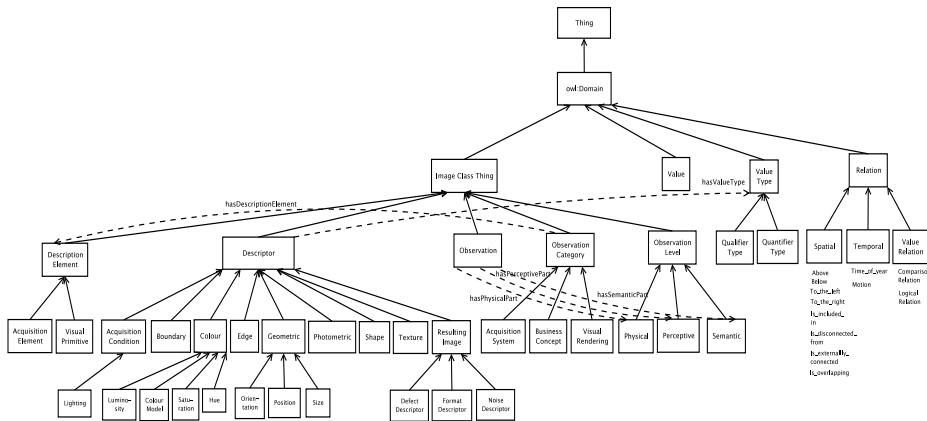


Fig. 2. Domain Ontology

### 3.3 Capability Ontology

The capability ontology (Fig. 3) contains the classes of video and image processing techniques. Each technique (or capability) is associated with one or more tool. A tool is a software component that can perform a video or image processing task independently, or a technique within an integrated vision library that may be invoked with given parameters. This ontology will be used directly by the Planner in order to identify the tools that will be used to solve the problem.

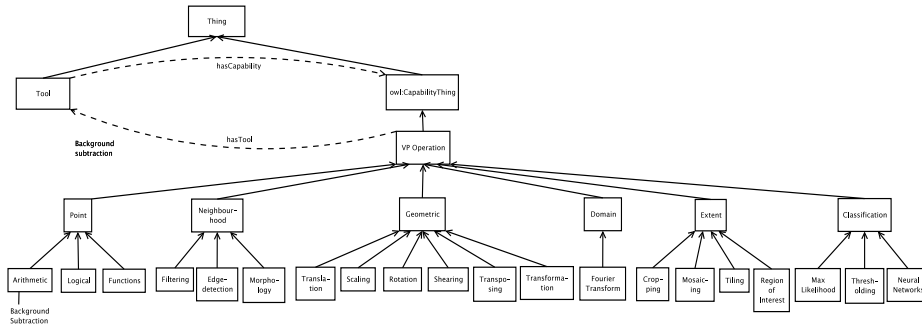


Fig. 3. Capability Ontology

## 4 Test Case

The ontologies formulated above have been applied to a real-world scenario in the EcoGrid [10] that utilises state-of-the-art technologies to establish a cyber-infrastructure for ecological research. This includes the integration of geographically distributed sensors, computing power and storage resources into a uniform and secure platform. Scientists can conduct data acquisition, data analysis and data sharing on this platform. One of the challenges is that a vast amount of raw



Fig. 4. Sample images extracted from EcoGrid video clips. From left to right: cluttered, partial object present, whole object present and non-presence of activity

data of varying qualities needs to be analysed efficiently and effectively (Fig. 4). Manual processing by ecologists, however, would be too time-consuming as a

minute’s video clip will take 15 minutes’ analysis time. This would not be feasible for data that amount to 1.86 Terabytes per year. Thus an automatic mechanism for processing these videos must be done in the most efficient manner. A full set of requirements for providing such a mechanism has been outlined in [11]. This includes an adaptive, flexible and generic architecture to allow for various video processing tasks to be conducted in an intelligent and optimal manner.

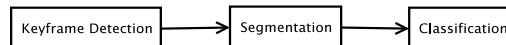
**Walkthrough.** Based on the devised ontologies, a walkthrough on how they are used to provide different levels of vocabulary for the users, vision tools and processes in a seamless and related manner is outlined here. The user, who is an ecologist, may have a high level goal such as “*Detect fishes in the videos*” in mind. This is represented and selected via the following selection-value pairs:

(Goal: Detection) [Detail Level: Occurrence = all\_occurrences]  
 [Performance Criteria: Processing Time = real\_time]

As a first step, the user interacts with the system by providing input values for the goal and constraints. The goal can be one of the goals specified in the goal ontology (Fig. 1). While the constraints are additional parameters to specify rules or restrictions that apply to the goal. These include qualifiers e.g. **Acceptable Errors**, **Optimisation Criteria**, **Detail Level** and **Quality Criteria** contained within the goal ontology.

Then the user describes the images to be processed. This description is given using the system which proposes descriptors contained in the domain ontology (Fig. 2). In our scenario, a description on the acquisition context and on the semantic content of the images is obtained; the lighting conditions changes very slowly, the camera is fixed (and also the background), images are degraded by a blocking effect due to the compression, and the fishes are regions whose colours are different from the background and are bigger than a minimal area (because when they are too small they are unusable). Thus, this goal is interpreted as “*The detection of all the occurrences of non-background regions on a fixed background.*”

As soon as the formulation of the user’s problem is made, a sequence of processes for execution using task decomposition will be sought by a Planning mechanism. Fig. 5 contains a breakdown of high-level tasks for the goal detection.

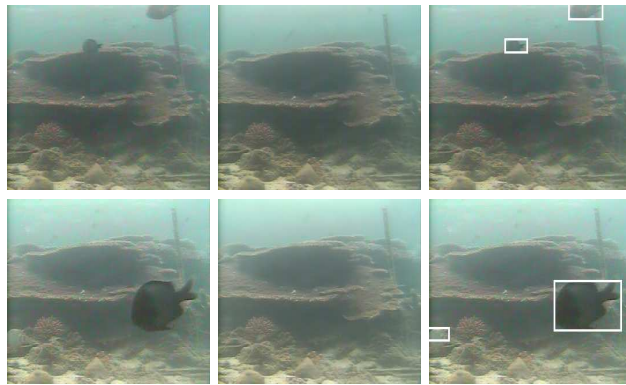


**Fig. 5.** High-Level Sequence of Process for Goal “Detect Fishes in the Videos”

Each task within this sequence could be further decomposed into sub-tasks. For instance, the ‘Segmentation’ task involves ‘Background Subtraction’ (under **Task/Process** in the goal ontology), which is done in three steps; background model construction, current frame and background model differencing, and background model update. All these sub-tasks are point arithmetic operations in the capability ontology. When a sub-task can no longer be decomposed, the tool or

technique identified for performing it can be applied directly on the video clip. The tool or technique available within the system is represented in the capability ontology. In the case where more than one tool is available to perform the sub-task, the tool with the best performing capability is selected.

In our case, the background model construction is done by averaging a series of successive images without any fish. Next, the difference between the current image and the background model is obtained. The model is updated using this current image to avoid future false detections due to the small changes of lighting conditions. False detections elimination (sub-task of the classification step) is achieved using a threshold on the minimal area given by the user. Fig. 6 shows two sample results obtained using the solution proposed by the modularised approach. As the detection is accurate, the same principles could be extended to achieve more complicated video processing tasks (such as motion analysis) as long as the ontologies capture the goals, domain descriptions and capabilities.



**Fig. 6.** Two Sample Results. From Left to Right for Each Row: Original Image, Background Model, Classified Result

## 5 Conclusion

The formulation of the video analysis problem description and solution could be tackled by modularising them using separate but inter-related ontologies. The three ontologies presented (goal, domain and capability) are extensive and describe the different aspects of video analysis in meaningful ways. This would prove to be useful as it facilitates a means to formalise the video analysis process and promotes reusability of applications. We have demonstrated the application of this modularised approach using an example from an ecological domain. The ontologies are high-level and general enough to be tailored towards building application ontologies by vision experts for solving tasks more specific to their problem domains. On a broader context, this effort is a significant step towards providing a semantically-enhanced framework in emerging infrastructures such as the Grid which would allow for real-time distributed image processing.

## References

1. Gomez-Perez, A., Fernandez-Lopez, M., Corcho, O.: *Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web*, 1st Edition. (2004)
2. Nouvel, A., Dalle, P.: An Interactive Approach For Image Ontology Definition. In: 13ème Congrès de Reconnaissance des Formes et Intelligence Artificielle, Angers, France (2002) 1023–1031
3. Maillot, N., Thonnat, M., Boucher, A.: Towards Ontology Based Cognitive Vision (Long Version). *Machine Vision and Applications* **16**(1) (2004) 33–40
4. Bombardier, V., Lhoste, P., Mazaud, C.: Modélisation et intégration de connaissances métier pour l'identification de défauts par règles linguistiques floues. *Traitement du Signal* **21**(3) (2004) 227–247
5. Hudelot, C.: Towards a Cognitive Vision Platform for Semantic Image Interpretation. Application to the Recognition of Biological Organisms. PhD thesis, Nice-Sophia Antipolis University (2005)
6. Town, C.: Ontological Inference for Image and Video Analysis. *Mach. Vision Appl.* **17**(2) (2006) 94–115
7. McGuinness, D., van Harmelen, F.: OWL Web Ontology Language. World Wide Web Consortium (W3C) (2004) <http://www.w3.org/TR/owl-features/>.
8. Matsuyama, T.: Expert Systems for Image Processing: Knowledge-Based Composition of Image Analysis Processes. *CVGIP* **48**(1) (1989) 22–49
9. Renouf, A., Clouard, R., Revenu, M.: How to Formulate Image Processing Applications ? In: *Proceedings of the International Conference on Computer Vision Systems*, Bielefeld, Germany (2007)
10. EcoGrid National Center for High Performance Computing, Taiwan: (<http://ecogrid.nchc.org.tw/>)
11. Nadarajan, G., Chen-Burger, Y.H., Malone, J.: Semantic-Based Workflow Composition for Video Processing in the Grid. In: *IEEE/WIC/ACM International Conference on Web Intelligence*. (2006) 161–165
12. Foster, I.: *The Grid 2 – Blueprint for a New Computing Infrastructure*. 2. ed. edn. Morgan Kaufmann (2004)
13. Liedtke, C., Blömer, A.: Architecture of the Knowledge Based Configuration System for Image Analysis "Conny". In: *ICPR'92*. (1992) 375–378
14. Clément, V., Thonnat, M.: A Knowledge-Based Approach to Integration of Image Procedures Processing. *CVGIP: Image Understanding* **57**(2) (1993) 166–184
15. Chien, S., Mortensen, H.: Automating Image Processing for Scientific Data Analysis of a large Image Database. *IEEE PAMI* **18**(8) (1996) 854–859
16. Clouard, R., Elmoataz, A., Porquet, C., Revenu, M.: Borg : A Knowledge-Based System for Automatic Generation of Image Processing Programs. *IEEE PAMI* **21**(2) (1999) 128–144
17. Draper, B., Hanson, A., Riseman, E.: Knowledge-directed vision : Control, learning, and integration. In: *Proc. of IEEE*. Volume 84. (1996) 1625–1681
18. Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, Y., Staab, S., Strintzis, M.G.: Semantic Annotation of Images and Videos for Multimedia Analysis. In: *ESWC*. Volume 3532 of LNCS., Springer (2005) 592–607
19. Renouf, A.: (Hermès - a human-machine interface for the formulation of image processing applications) <http://www.greyc.ensicaen.fr/~arenouf/Hermes>.