Issues in representing context illustrated by scene interpretation applications

François Brémond and Monique Thonnat
I.N.R.I.A., 2004 rte des Lucioles - BP 93
06902 Sophia-Antipolis Cedex, France
Tel : +33 93 65 76 59, Fax : +33 93 65 79 39
Email : francois.bremond@sophia.inria.fr

Keywords: Context modeling, knowledge representation, knowledge-based system, scene interpretation, view-point.

Abstract

This paper tackles several issues of the representation of context in knowledge-based systems.

First we propose a context definition through the description of the different information types manipulated by a process. Thanks to this definition we explain the role of the granularity level of processing and the role of the abstraction level of application in modeling context. Then two rules are provided helping with the construction of context representation. We illustrate the proposed context model with the example of the scene interpretation process. Then we present a representation of its context. Through this specific representation two more general issues of context representation are tackled : how context can be built and organized and how its contents can be re-used for other applications. In the last part of the paper we propose solutions to deal with these issues. Thus the proposed context representation model allows us to see the representation under different view-points. At the same time it allows us to gather all contextual information in one place. As the representation is partially generic it also makes possible the re-use of context contents for different interpretation applications.

1 Introduction

The design of systems which interpret a scene described by image sequences is an increasing research field. Although they use efficient algorithms, taking part in the interpretation process, the results of the overall system are not completely satisfactory. Several teams have developed approaches to overcome this problem. For example, we can quote H. Nagel [7], J. Schirra [8], A. Bobick [1] and the esprit VIEWS project (Visual Inspection and Evaluation of Wide-area Scenes) [4]. This work can be related to other domains as case-based reasoning [11].

In [2] we explained why context is particularly useful for interpretation systems and we described a way of representing it. Taking advantage of the specific case of scene interpretation applications, this paper tackles general issues of context representation in knowledge-based systems : what is context, how context can be built and organized and how its contents can be also re-used for other applications.

Firstly we present a context model through the description of the different information types manipulated by process. Then based on this model we briefly recall how we represent the context of the interpretation process. Finally we present solutions to help with the construction of context representation and to help to re-use its contents for different interpretation applications.

2 Definition

2.1 Contextual information

The definition of the context of a process depends on the process nature. For H. Nagel [7] the context of an action analysis process is a complex structure, comprising generic descriptions for spatial structures, temporal changes associated with these structures and the intention aspect of the action. For T. Strat [10], the context of an image understanding process is, in the broadest sense, any and all information that may influence the way a scene is perceived. More generally, a process uses three types of information : knowledge, contextual information and factual information. **Knowledge** is always valid. It is directly connected with the process goals and often belongs to a well defined model. If the process misses one part of knowledge, it is no longer able to compute results. **Contextual information** depends on applications, but it remains constant during processing. It is an accessory information, but it may become essential to handle particular situations and enables to improve the processing. It is all the additional information needed by the process to work efficiently. **Factual information** depends on the processing states. Its life of time is often short. It corresponds to input data and computed data. Thus we propose to define the contextual information of a process as the information verifying two conditions :

- its value remains constant during processing,
- its value changes when the process is used for another application. (1)

This definition of context has two main consequences. First we have to choose from which granularity level of processing we consider an information as factual rather than contextual. Second we have to choose from which abstraction level of application we consider an information as contextual rather than belonging to knowledge.

We illustrate these two choices through the example of the scene interpretation process. So in the next section we start to choose its granularity level of processing.

2.2 Several granularity levels of processing

The class of applications we are interested in, is the automatic interpretation of indoor and outdoor partially structured scenes with a fixed monocular color camera. Given image sequences describing a scene, an interpretation system has to identify the behaviors of mobile objects. In our case, mobile objects can be either humans or vehicles. They are the moving scene objects, as opposed to static objects that belong to the static environment. By interpretation, we mean the overall interpretation process from image processing to behavior analysis. This process can be divided in three main tasks, which can overlap each other: **image processing**, **mobile object recognition** and **behavior analysis**. From images, the image processing task detects moving regions. Then the mobile object recognition task associates these regions to form and to track mobile objects. Finally, the behavior analysis task interprets event occurrences concerning mobile objects and computes their behaviors. In addition the object recognition and behavior analysis tasks both include **spatial reasoning**, that can be seen as a basic task. Thus four specific tasks compose the interpretation process, which are the three main interpretation tasks and the spatial reasoning task. So for interpretation systems two levels of processing can be used: the global process or the four tasks. We choose to delimit context according to the tasks, because they belong to different well defined knowledge domains. Using definition 1 for the context of an interpretation task, we define the context of the global interpretation process as following (*task-i* denotes any task of the interpretation process):

$$\text{context[global process]} = \bigcup_i \text{context[task-i]} \tag{2}$$

This definition provides us with a rule that can decide whether a piece of information verifies the first condition of definition 1: if it remains constant during the processing of at least one task that uses it, then the first condition is verified.

2.3 Several abstraction levels of application

As a second consequence of definition 1, we have to choose the abstraction level of application in order to decide on which information domains the process depends. For the example of the interpretation process, we can consider an application at the general level of video-surveillance, at the more specific level of metro station surveillance or at the even more specific level of the surveillance of a specific station observed by a specific camera.

The definition of interpretation process context depends on the chosen abstraction level of application. In this paper the surveillance of a specific metro station is considered as an application. So now we can take inventory of the sources (as suggested in [10]), where contextual information may come from :

- **Scene Environment information** - It embraces spatial structures (e.g. tessellations, calibration plans), static objects (e.g. pillars, escalators), optical characteristics (e.g. reflections on the ground, occlusions) and behavior characteristics (e.g. exit and cluttered zones, roads).
- **Image Acquisition information** - It includes camera characteristics (e.g. camera model, focal length, color, filter), image characteristics (e.g. image size and type, date and time of acquisition) and acquisition characteristics (e.g. camera orientation and camera position).
- **Derived Temporal information** - It is obtained as a result of earlier executions of interpretation tasks. This information can be seen as the accumulated information about the past and the prediction about the future (e.g. already detected and identified mobile objects).
- **User Request information** - In an interactive way, a human operator can provide context to the overall system during processing. A typical example for surveillance system is an operator request for detecting specific persons.

The interpretation process depends on all these source domains of contextual information. This property provides us with a rule that can decide whether a piece of information verifies the second condition of definition 1 : if it belongs to one of these source domains, then the second condition is verified.

3 Context representation

3.1 Space representation

This section presents a global context representation for the four tasks of the interpretation process. As this representation and its utilization were already presented in [2], this article gives a brief description. For each task of the interpretation process a piece of context is represented by a frame called a **context element**. Using the frame formalism a context element is defined as a set of slots corresponding to the properties of the piece of context. As the interpretation process context is mainly composed of scene environment information and as it is suggested in [6], we organize context around the space representation. In a first time this section describes the representation of the space and then explains how the context of each interpretation task is linked to it.

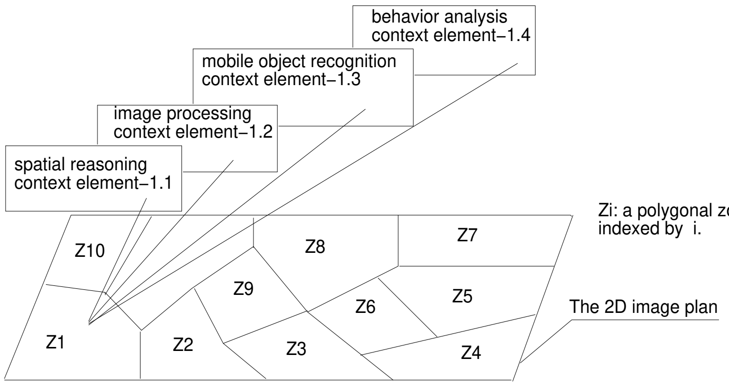

Figure 1: A context element for each interpretation task is linked to a polygonal zone of the decomposition of space.

For an interpretation system the space corresponds to the projection of the scene on an image. We decide to represent the space by a decomposition of the 2D image plan into a partition of zones delimited by polygons. Polygons have been chosen because they are structures simple enough to be easily implemented, and flexible enough to match human knowledge. The polygonal zones have to be drawn off line by a human operator. This drawing operation is a necessary stage before using the interpretation system. Each polygonal zone is linked to four context elements, one element for every interpretation task. A zone delimits the place where its corresponding context elements are useful. Figure 1 shows the polygonal zone *"Z1"* with its four corresponding context elements ranged from *"spatial-reasoning-context-element-1.1"* to *"behavior-analysis-context-element-1.4"*. For example the object recognition task can use the *"mobile-object-recognition-context-element-1.3"* to improve its recognition of objects moving in the zone *"Z1"* during all the processing of the image sequence. Therefore the **decomposition of space** is the support of context representation and gathers all its contents.

3.2 Context representation for the interpretation tasks

The interpretation context is represented through a set of context elements. This section gives some examples of their type and contents for each interpretation task.

Context in spatial reasoning :
The context element of the spatial reasoning task is used to compute 3D spatial relations on mobile objects. This context holds the description of a calibration matrix. For any point belonging to the polygonal zone linked to the context element, the matrix allows the computation of its 3D scene coordinates depending on its 2D image plan coordinates. So the spatial reasoning context element is made up of three slots : the calibration matrix, the coefficient giving the accuracy of the scene coordinates computation, and the link to its corresponding zone.

Context in image processing :
The context element of the image processing task is generally used to adapt image processing programs to new scene conditions and to focus the system attention to specific image regions of interest. As in our interpretation system we use image processing programs that do not handle scene characteristics, we define the same image processing context element for all the polygonal zones of the decomposition. This context element contains global slots like the size of one image and the frequency of the image sequence.

Context in mobile object recognition :
The context element of the mobile object recognition task is used to handle specific problems. This task may lose tracks of mobile objects due to optical irregularities, such as reflections on the ground, shadows, cluttered zones, blinking lights, occlusions. For this reason the context element of the mobile object recognition contains mainly three slots : the optical irregularity descriptions, the uncertainty coefficient of the recognized objects and the link to its corresponding polygonal zone. This zone is supposed to delimit the optical irregularity influence. This representation implies that a library of optical irregularities is already defined in the interpretation system.

Context in behavior analysis :
The role of the behavior analysis task is to match the numerical properties of mobile objects with behavior models expressed in natural language. Because of the gap between these two notions [9], only simple behavior models can be used. The context element of the behavior analysis is defined as a bridge between object properties and complex behavior models. As underlined by Howarth [5], most areas are named either by their location (e.g. a kitchen) or by the actions that occur in them (e.g. cooking), providing to the analysis task

important clues on expected behaviors. The behavior analysis context element holds four slots: the static objects of the environment belonging to the area (e.g. a seat), properties on expected mobile objects (e.g. the speed limit is low), expected behavior models (e.g. to seat down) and the link to its corresponding polygonal zone. As remarked before, this context element description implies that libraries of static objects, of object properties and of behavior models already exist. These libraries are predefined in the interpretation system.

Therefore for each interpretation task, all context elements are represented with the same formalism and are linked to the decomposition of space.

4 Discussion

When we try to represent the context of a process we often have to deal with two general issues: how context can be built and organized and how its contents can be re-used for other applications. In interpretation applications these two issues have to be tackled. This section explains the solutions chosen for the presented representation of context.

4.1 Handling several contextual information domains

A first problem arises when the domain of a process context is defined as the union of several task context domains. By definition context depends on applications and so its contents have to be acquired before its utilization. In our system contextual information is provided by human operators. Since four task context domains are defined for the interpretation process, the context acquisition requires several operators, at least one for each interpretation task. So a first problem is the mix of all the acquisitions of task context domain. This problem is crucial when large systems are developed. The solution we propose, is to use a multi view-points representation based on the frame formalism. In this representation a view-point is a way of seeing context for a specific interpretation task by filtering the corresponding context element slots. For interpretation systems four view-points are needed, one for each interpretation task. Thus view-points divide contextual information and help operators with its construction.

A second problem comes from the choice of definition 2 for the interpretation process context. Because of definition 2 we consider that all the context of a task, even provided during the processing of another task, belongs to the global process context. This case arises with the source domain of derived temporal information. For example the degree of interest of a mobile object is computed by the mobile object recognition task but it is also useful for the image processing task to select on the next image the regions of interest. So the image processing task needs this information to improve its computation. Therefore a mechanism is needed to share the common information between interpretation tasks. The solution we choose is to link the common information to the decomposition of space. To solve the above example we have linked a list of mobile objects of interest to each polygonal zone. The list is computed by the mobile object recognition task and used by the image processing task. This solution is possible only because in interpretation applications the decomposition of space is a basic structure common to all tasks. For this reason the decomposition of space has been chosen to gather all context domains in one place. It can be seen as a reasoning support of the interpretation process and helps with the organization of the context representation.

Therefore the context representation allows us to divide context during its construction phase through view-points and to share context during the processing phase through the decomposition of space.

4.2 Handling several source domains of contextual information

The issue of re-using old contextual information is all the more relevant that a high abstraction level of application is considered. If a high abstraction level is chosen, then the process depends on several source domains of contextual information, involving all these domains in the process context and making its contents numerous.

In interpretation applications context depends on four source domains of contextual information. Thus a big amount of contextual information has to be acquired, especially because of the scene environment information. A solution is to automate the acquisition phase, but this issue is particularly difficult when information on human activities has to be acquired. The solution we choose is to re-use the contextual information of old applications. As suggested in [3] a main issue in context formalization is to abstract information. In our representation this idea is realized by the utilization of symbolic information. Instead of representing context with a complete description, only symbolic information is used. The complete description is predefined in libraries belonging to the interpretation system. For example for the behavior analysis context just symbols are used for the description of the static objects of the environment and for the description of behavior models. Thus one part of the interpretation context is generic and can be used for different applications (e.g. for the surveillance of different metro stations).

Despite the utilization of symbolic information context acquisition is still a tiresome phase. For this reason we have developed a graphical user interface, which acquires contextual information interactively with an operator and builds the context representation of a scene. It is implemented in C language and uses Motif toolkit [2]. Figure 2 shows polygonal zones drawn by a human operator and a context element linked to the decomposition of

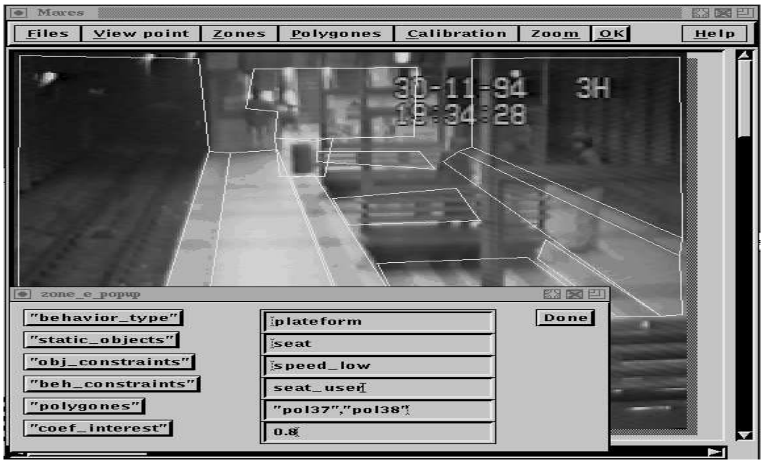

Figure 2: A human operator is using the interface to fill the values of the object recognition context element for the polygonal zones on the ground (polygonal zones are drawn in white).

space.

5 Conclusion

This work proposes a general definition of context for knowledge-based systems through the description of different types of information. This definition depends on the granularity level of the processing and on the abstraction level of the application. The context definition allows us to define the context of a process composed of several tasks. It allows us also to bring to light the role of the source domains of contextual information. Deduced from the context definition, two rules are presented to build context domain. Then, two general issues in context representation are tackled : how context can be built and organized and how its contents can be re-used for other applications.

To handle these issues we propose in one hand to divide context thanks to a multi view-points representation based on the frame formalism and in the other hand to gather the whole context in one place through the decomposition of space. So the presented representation allows us to see context under different view-points and at the same time to share common contextual information between all interpretation tasks. We propose also a way to abstract contextual information in order to obtain a partially generic context representation and to make possible the re-use of its contents. Thus context representation becomes easier to build, to acquire and to use. Our future works will try to generalize these solutions and to adapt them to other types of applications.

References

[1] A. Bobick and C. Pinharez. Using approximate models as source of contextual information for vision processing. In *IEEE workshop on Context-Based Reasoning (ICCV'95)*, April 1995.

[2] F. Brémond and M. Thonnat. A context representation for surveillance systems. In *European Conference on Computer Vision, workshop on Conceptual Descriptions from Images*, Cambridge, April 1996.

[3] J. Mac Carthy. Notes on formalizing context. In *IJCAI, Chambery, France*, pages 555–560, 1993.

[4] D. Corrall. Deliverable 3: Visual monitoring and surveillance of wide-area outdoor scenes. Technical report, EP 2152: VIEWS, June 1992.

[5] R. Howarth. Interpreting a dynamic and uncertain world: high-level vision. *Artificial Intelligence*, 9(1):37–63, February 1995.

[6] M. Mohnhaupt and B. Neumann. Understanding object motion: Recognition, learning and spatiotemporal reasoning. Research Report FBI-HH-B-145/90, University of Hamburg, 1990.

[7] H. H. Nagel. From image sequences towards conceptual descriptions. *Image and Vision Computing*, 6(2):59–74, May 1988.

[8] J. R. J. Schirra. A contribution to reference semantics of spatial prepositions: The visualization problem and its solution in vitra. Research Report 75, VITRA-report, December 1990.

[9] R. Srihari. Computational models for integrating linguistic and visual information: A survey. *AIR*, 8(5-6):349–369, 1994.

[10] T. Strat. Employing contextual information in computer vision. In *DARPA93*, pages 217–229, 1993.

[11] R. Turner. Context-sensitive, adaptive reasoning for intelligent AUV control: Orca project update. In *proc. of the 9th Int'l Symposium on Unmanned Untethered Submersible Technology (AUV'95)*, New Hampshire, 1995.