How Useful Is Formal Knowledge Representation for Image Interpretation?

Bernd Neumann and Carsten Schröder

Abstract

In this position paper we provide arguments for the following main points:

(1) Formal knowledge representation and well-founded inference processes can help to clarify, unify and support the construction of image interpretation systems.

(2) Today's terminological systems (like LOOM) fail to support basic image interpretation inference strategies, in particular hypothesis generation and spatial reasoning.

(3) Terminological systems can be extended to provide inference services useful for image interpretation tasks.

1. Introduction

The topic of this contribution is formal knowledge representation and its use for image interpretation. We do not raise a discussion about any kind of knowledge representation - there can be no doubt about the need of explicit knowledge representation in knowledge-heavy interpretation tasks. Our point of discussion is the usefulness of representing knowledge in a formally accountable way, typically based on some kind of logics. In particular we look at terminological systems of the KL-ONE family like CLASSIC and LOOM which are available world-wide and have been employed for commercial applications.

It is a well-known phenomenon in the AI world that logic-based knowledge representation seems to be a prerequisite for successful AI research but dispensable for most practical AI applications. On a first glance this is understandable, as research is mainly concerned with analysability whereas applications are mainly concerned with performance. Formal knowledge representation has indisputable merits for analysing the formal structure of a problem and its solution, but is likely to be the wrong tool for implementing practical systems.

But it is important to consider a knowledge representation formalism not only for its analytical potential but also for the inference services which may become available. Terminological systems like LOOM offer a variety of such services including inheritance, subsumption, concept and object classification, consistency and others. A powerful terminological classifier, for example, could provide the correct classification of an unknown object in an image interpretation task based on specific attributes determined from the image and on the conceptual descriptions of object classes in the knowledge base.

Could it in deed? One of the purposes of this contribution is to illuminate the usefulness of such tools for image interpretation. In Section 2, we discuss experiences with a change interpretation task in aerial images. Using the object classifier of LOOM, interesting changes like a runway elongation can be automatically determined from basic observations in segmented images.

However, this approach is inherently limited because of its purely deductive nature.

In Section 3 we consider the use of terminological systems in support of non-deductive reasoning strategies, in particular of the classical hypothesise-and-test paradigm of image interpretation. The main service we like to get from the knowledge base is a good set of candidate concepts for the classification of an unknown object. Several extensions of current terminological systems are suggested, including steps towards a probabilistic extension of description logics.

Our conclusions are presented in Section 4. On the positive side, we believe that formal knowledge representation can provide the basis for powerful inference mechanisms adapted to image interpretation requirements. Exploiting such inference mechanisms may lead to image interpretation systems with improved properties regarding correctness, ease of development and reusability. On the negative side, tools providing such inference services are not yet available and further research is still required to this end. Complexity barriers may still prevent such tools to become useful for practical applications.

2. Formal Classification

One of the most powerful inference services offered by current terminological systems is individual classification. This is a process which determines the most specific superordinate concepts of a knowledge base for an unknown object described by attributes and relations. For example, if a runway in an aerial image is conceptually defined as a rectangular road with a certain minimal length and width and certain connections to taxiways, a classifier should be able to determine whether this conceptual description is satisfied by some piece of aerial imagery.

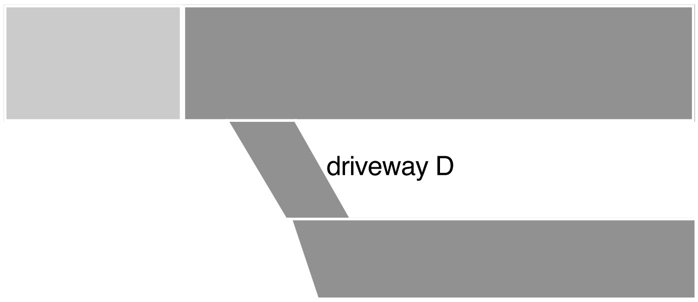
In view of the complexity of hand-coded classification processes, it would certainly be an advantage to make use of a classifier offered as an inference service of a terminological system. This would not only save software development efforts but would also allow statements about correctness and completeness. Furthermore, making use of the knowledge representation tools, one would hope that the formal semantics of the terminological system would facilitate knowledge reuse, for example through the use of ontologies. We therefore decided to explore the usefulness of a terminological system for a concrete task, change interpretation in aerial images.

The problem is related to disarmament surveillance where certain changes of certain objects (e.g. a runway elongation) are interesting and should be recognised. In our experiment, two aerial images of the same area but taken at different times are assumed to be available. Finding interesting changes amounts to recognising corresponding objects in the two views and classifying possible differences.

Employing a terminological system, its first role is to provide conceptual descriptions of the object classes of the domain and give precise definitions of the changes which are to be recognised. Its second role is to automatically classify a concrete image pair according to the conceptual definitions. To this end, segmentations in terms of regions and geometrical properties are provided which can be taken as primitives for higher-level descriptions.

[Lange and Schröder 94] provide a detailed report of the experiment. In this section, we can only present some of the material. Our main purpose is to illuminate problems with current terminological systems when applied to image interpretation.

At the core of the system are conceptual descriptions to recognise runway structures and possible changes as sketched below.



Structure of runway, driveway and taxiway

For a runway to be automatically recognisable, a runway concept is defined in terms of certain sufficient (and necessary) conditions. The definition starts with geometric constraints for the shape of the runway (using shape concepts defined elsewhere). Then it is required that there is at least one filler of the role has-connecting-driveway, and all the fillers must satisfy simple constraints for the length and the width of a driveway. Note that these fillers are not required to be of the concept driveway. After this follows a rather complex formulation of relative constraints for the driveway and taxiway associated with the runway. For a runway (named ?x) it is required that there exists at least one filler of the role has-connecting-driveway (named ?y) which satisfies the geometric constraints for a driveway with respect to the runway ?x and which has at least one neighbour (named ?z) which satisfies the geometric constraints for a taxiway with respect to the runway ?x and the driveway ?y. The conjunction of all these conditions is necessary as well as sufficient for an object to be a runway.

```
(defconcept runway
   :is
   (:and
     roadlike-structure
     rectangle
     (:all has-length1 (:through 2150 4000))
     (>= has-width 45)
     (:at-least 1 has-connecting-driveway)
     (:all has-connecting-driveway
        (:and  (= has-length1 has-length2)
               (>= has-width 23)))
     (:satisfies (
        (?x)
        (:about ?x
           (:at-least 1
              has-connecting-driveway
              (:satisfies (
              (?y)
              (:and
                 (:predcall satisfies-driveway-constraints ?y ?x)
                 (:about ?y
                    (:at-least 1
                       has-neighbour
                       (:satisfies (
```

```
                    (?z)
                    (:predcall
                      satisfies-taxiway-constraints
                      ?z ?y ?x)))))))))))))
```

The role has-connecting-driveway plays a key part in the runway concept. It is defined as follows:

```
(defrelation has-connecting-driveway
  :is
  (:and  has-neighbour
         (:domain roadlike-structure)
         (:range (:and roadlike-structure
                       (:at-least 2 has-neighbour
                                    roadlike-structure)))))
```

Two objects x and y are in the relation has-connecting-driveway if and only if they are roadlike-structures which satisfy the relation has-neighbour, and y satisfies the relation with at least two roadlike-structures. The role has-neighbour is meant to be the relation of topological adjacency. In order to provide efficient computation of objects adjacent to one another, a definition is provided in terms of a function outside the role hierarchy:

```
(defrelation has-neighbour
  :function ((x) (compute-neighbouring-objects x))
  :characteristics (:symmetric :multiple-valued))
```

The example definitions given above allow to follow the basic reasoning path of an object classifier in a terminological system when recognising a runway. Let us assume that a segmentation of a multispectral image of an airport is available which contains, among others, three regions corresponding to a runway, a driveway and a taxiway, respectively. The regions, initially assigned to the general concept scene-object, can be classified as roadlike-structures because of material and shape properties. At this point the conditions for a runway will be checked by the object classifier. Eventually, this leads to checking the has-connecting-driveway constraint for region R (the runway candidate). By definition, has-neighbour is checked and the generator function compute-neighbouring-objects is called to produce role fillers including the driveway region D, among others. D in turn must fulfil the has-neighbour relation with at least two partners, hence the generator is called again, delivering R and T, among others. Finally, the complex :satisfies term must be checked. It requires that a runway candidate x satisfies detailed geometrical constraints (defined elsewhere) relative to a driveway candidate y and a taxiway candidate z. Again, the generator function compute-neighbouring-objects is called to generate topologically adjacent candidates. If R, D and T fulfil the constraints, all conditions are met and R is classified as a runway.

We show now how a runway elongation can be conceptually defined to be automatically recognisable by a classifier. First we define a change as any two scene objects at two different times:

```
(defconcept basic-change
  :implies
  (:and  (:exactly 1 has-pre)
         (:exactly 1 has-post)
         (>   (:compose has-post has-time)
              (:compose has-pre has-time))))
```

The roles has-pre and has-post (definitions not shown) refer to two time slices with ordered time index. An elongation is a special kind of change defined as follows:

```
(defconcept elongation
  :is
  (:and  basic-change
         (:relates has-contained-object has-pre has-post)
         (> (:compose has-post has-length1)
              (:compose has-pre has-length1)
         (> (:compose has-post has-length2)
              (:compose has-pre has-length2)
         (= (:compose has-post has-width)
              (:compose has-pre has-width))))
```

The scene object in the first time slice must be spatially contained in the latter (definition not shown), and certain length and width inequalities must be fulfilled. Using this general elongation concept we can finally define a runway elongation:

```
(defconcept runway-elongation
  :is
  (:and  elongation
         (:all has-pre runway)
         (:all has-post runway)))
```

Two scene-objects at different times meeting the conditions spelled out in the concept definitions can now automatically be classified as an instance of a runway-elongation. At first a basic-change instance has to be created, then this instance is automatically specialised into an elongation, if the conditions are met. The relation has-contained-object is evaluated on demand by the object classifier.

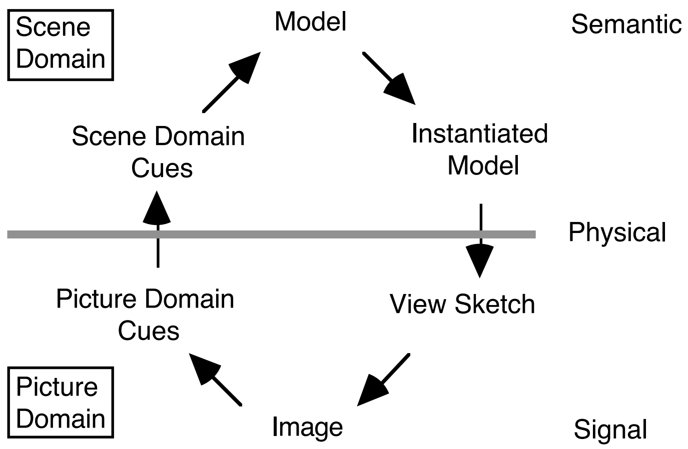
Those who have not yet used a terminological system will probably need time to get used to the complicated expressions. On the one hand, they provide a precise account of what we want to recognise. Assumptions which would remain implicit in informal definitions, have to be spelled out, for example about the uniqueness of role fillers or the sufficiency of conditions. On the other hand, the clarity and the possible scope of concept definitions suffers from a lack of expressiveness of terminological languages. This is generally recognised as a shortcoming, but the price for more expressiveness is heavy: Reasoning processes may become computationally more complex (intractable) or even undecidable [Nebel 90]. Note that by this reason the description language of LOOM which we have used in this experiment is undecidable, resulting in incomplete reasoning services.

One of the limitations of formal knowledge representation which tend to make life difficult for practitioners concerns aggregations. See, for example, [Matsuyama and Hwang 90] for a discussion of related problems. In our case, an aggregate of a runway, a driveway and a taxiway has to be recognised - let us call it an airfield. A natural way of defining an airfield would be in terms of parts - roadlike structures - and additional constraints between the parts. However, the formal semantics of parts and wholes is problematic. To put it simple, it is difficult to express that parts become something special when they constitute an aggregate. Hence terminological systems do not offer a predefined "part-of" role like many frame systems. From a more practical point of view, one of the problems is to induce a classifier to assemble suitable parts into an aggregate. Different from the usual strategy where classification advances along the specialisation hierarchy, an aggregate should be formed from parts.

Another problem concerns the treatment of space and time. Almost all terminological systems have no built-in primitives to support spatial or temporal reasoning. For example, there is no efficient access to spatially adjacent objects unless one provides user-defined generator functions (see our example). Such functions, however, are opaque and cannot be placed into the role hierarchy (apart of a crude characterisation). If, for example, a second function for a more restricted neighbourhood access would be defined, the subsumption relation between the two functions could not be exploited and spatial reasoning would remain incomplete.

The most serious problem of using a classifier concerns the control structure of the image interpretation process. It is useful at this point to remember the basic hypothesise-and-test paradigm which reflects our understanding of image interpretation control (see the diagram of Kanade 78 below). An essential ingredient of efficient image interpretation is hypothesis formation from partial evidence (bottom-up) followed by hypothesis verification (top-down). The efficiency gain is twofold: (i) Making good guesses reduces bottom-up search, and (ii) focusing on missing evidence reduces low-level computations.

A formal classifier does not conform with this paradigm. In particular, there is no hypothesis generation, no guessing of likely classifications, not even a computation of possible classifications. Classifications are deduced from evidence which must be completely provided beforehand. There is no mechanism to compute missing evidence as it is needed. Of course, terminological systems may offer hooks to influence control to some degree. But a formal classifier is designed as a deductive service and not as an image interpretation control regime.



A Model for Image Understanding
(adapted from Kanade 78)

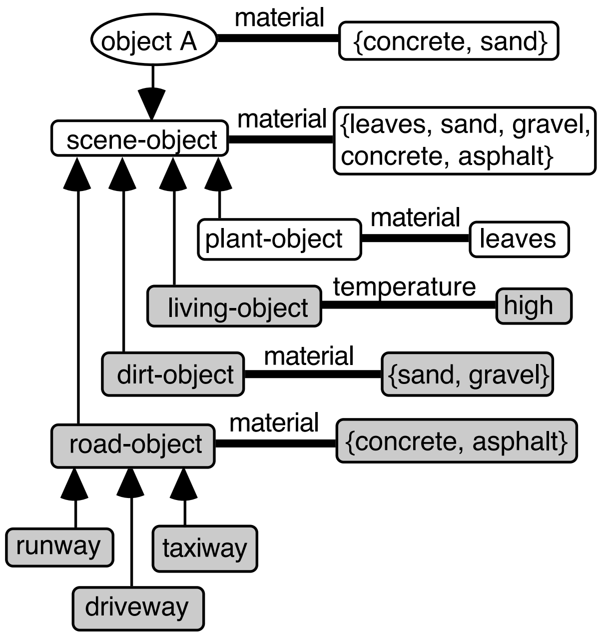
This does not mean that a terminological system is basically useless for image interpretation. We still have the advantage of precise conceptual definitions with well-defined semantics. But we need reasoning services which are better adapted to image interpretation tasks. For example, we would like to support hypothesis generation and hypothesis verification. In the next section we shall propose some extensions of current terminological systems which may be useful for this purpose.

3. Supporting Hypothesis Generation with a Terminological System

How can a terminological system support the basic hypothesise-and-test cycles of image interpretation? The reasoning services which we propose now are fairly modest extensions of current terminological systems. They do not require basically new functionality. We suggest to perform image interpretation in a basic hypothesise-and-test framework where hypotheses are generated from partial (inconclusive) evidence and further evidence is computed on demand to verify or falsify hypotheses. This can take place at any level above the segmented image. The main difference to using a formal classifier as shown in the previous section is the expected efficiency gain due to the top-down computation of evidence. Logically, classification remains a deductive process.

At this point it is worth noting that different logical frameworks have been proposed where image interpretation is modeled as abduction [Matsuyama and Hwang 90] or as a logical model construction process [Reiter and Mackworth 87]. The discussion of these approaches and of ways to support them with a terminological system is beyond the scope of this paper.

To support a hypothesise-and-test strategy, a useful service of a terminological system could be consistent-hypotheses-generation. Given partial evidence about an object, what are the remaining concepts into which this object could possibly be classified? Different from a conventional object classifier which generates all deducible concepts, consistent-hypotheses-generation would generate all non-refutable concepts. This can be a useful service as the full power of terminological definitions can be exploited to avoid further elaboration of doomed hypotheses.



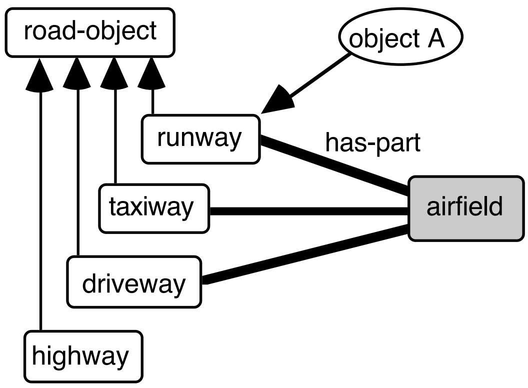
Hypotheses consistent with partial evidence

On the other hand, consistency with evidence is a rather weak criterion for a hypothesis to be promising as the set of consistent hypotheses also includes those for which there is no evidence whatsoever. Consider the situation depicted above where the material properties concrete or gravel have been determined for an unknown object A, so far classified as scene-object. The set of consistent hypotheses comprises the shaded cells and includes living-object for which no evidence is yet available.

A frequently used reasoning technique in image interpretation is part-whole-reasoning. Given evidence about parts, hypothesise and try to verify the whole. A part-whole relationship expresses some sort of aggregation, spatial, temporal or otherwise. By exploiting part-whole relationships we assume that evidence for a part somehow increases the likelihood that the aggregate may be present.

As mentioned earlier, current terminological systems have no predefined part-of semantics nor any notion of likelihood. Nevertheless, part-whole relationships can be expressed in terms of roles connecting aggregates and parts. In order to support part-whole reasoning based on such structures we propose an inference service which provides access to aggregate concepts of which an object might be a part. The idea is illustrated below. Given partial evidence in terms of an object A which is classified as a runway, the terminological system proposes an airfield as an interesting hypothesis since the runway is a possible filler of the has-part role of the airfield.

While this does not resolve the subtleties of part-of semantics, it still seems to be a generally useful inference service since it exploits conceptual information about the roles which some object may possibly play. If an object participates in many part-of relationships, i.e. if the role is 1-to-many, there are many aggregate candidates and the inference power may be small just as in situations where partial evidence is inconclusive. But if the role singles out only one aggregate, this information may speed up further processing considerably.



Aggregate consistent with partial evidence

The usefulness of hypotheses-generating inference services can be further improved by providing a ranking between hypotheses by means of a measure of likelihood. Carried out with all consequences, this would be equivalent to providing a probabilistic framework for classification. How this can be done in a manner consistent with a terminological system, is a topic of ongoing research [Jaeger 94].

As a less ambitious step into this direction we suggest to exploit A-box statistics. The A-box of a terminological system contains assertions about individual objects, in this case about the objects encountered in concrete image interpretation tasks. In the absence of other sources of statistical information, A-box statistics provide a valid basis for generating expectations about the contents of new images. In particular, given partial evidence for an unknown object, a likelihood ranking of candidate hypotheses can be provided based on past experience.

Note that hypotheses ranking is nothing out of the ordinary for any odd image interpretation system. By suggesting this service for image interpretation in a terminological framework, we try to provide minimal services which such a framework must offer to be competitive.

4. Conclusions

In this contribution we have taken a close look at terminological systems and their use for image interpretation. Illustrated by an extended example taken from [Lange and Schröder 94] we have shown that fairly complex changes in aerial images can be precisely defined and automatically recognised using the object classifier of the terminological system. Thus, in principle, the construction of image interpretation systems can be greatly facilitated by employing standardised knowledge representation and reasoning services instead of specially developed knowledge structures and interpretation algorithms.

The example has also shown, however, that an object classifier performs straight bottom-up deduction. A hypothesise-and-test control structure which is an important ingredient for efficient image interpretation, is not supported. In order to adapt terminological systems to the needs of image interpretation, mechanisms to generate and administer hypotheses have to be provided. Several modest extensions of the reasoning services of a terminological system have been proposed to this end.

At the bottom line, we feel that the advantages which can be gained from formal knowledge representation and reasoning outweigh the problems, in particular when we think of the need to develop complex knowledge-heavy image interpretation systems with maximal reuse of standard components.

References

Jaeger 94
Probabilistic Reasoning in Terminological Logics
M. Jaeger
Proc. KR-94, Morgan Kaufman, 1994.

Kanade 78
Region Segmentation: Signal vs. Semantics
T. Kanade
Proc. IJCPR-78, 1978, pp. 95-105.

Lange and Schröder 94
Analysis and Interpretation of Changes in Aerial Images: Knowledge Representation and Spatial Reasoning
H. Lange, C. Schröder
ISPRS Comm. III Symp. 94, International Archives of Photogrammetry and Remote Sensing, Vol. 30, Part 3/2, pp. 475-482.

Matsuyama and Hwang 90
SIGMA - A Knowledge-Based Aerial Image Understanding System
T. Matsuyama, V.S. Hwang
Advances in Computer Vision and Machine Intelligence, Plenum, 1990.

Nebel 90
Reasoning and Revision in Hybrid Representation Systems

B. Nebel
Lecture Notes in Artificial Intelligence, Vol. 422, Springer, 1990.

Reiter and Mackworth 87
The Logic of Depiction
R. Reiter, A. Mackworth
TR 87-23, Dept. Computer Science, Univ. of British Columbia, Vancouver, Canada, 1987.