NETHERLANDS GEODETIC COMMISSION

PUBLICATIONS ON GEODESY

NEW SERIES

ISSN 0165 1706

NUMBER 44

KNOWLEDGE-BASED INTERPRETATION OF AERIAL IMAGES FOR UPDATING OF ROAD MAPS

MARLIES DE GUNST

1996

NEDERLANDSE COMMISSIE VOOR GEODESIE, THIJSSEWEG 11, 2629 JA DELFT, THE NETHERLANDS
TEL. (31)-(0)15-2782819, FAX (31)-(0)15-2782745

This publication is accomplished with the support of the Faculty of Geodetic Engineering, TU Delft
Printed by Meinema B.V., Delft, The Netherlands

CONTENTS

CHAPTER 1

INTRODUCTION

In the photogrammetric practice there is a need to automatize acquisition of topographic information from aerial photographs. However, especially tasks involving interpretation capabilities of human operators are hard to automate. Digital photogrammetry can benefit from experiences with knowledge-based concepts in computer vision. Within this field our goal is to investigate the potential of knowledge-based image processing techniques for interpretation of aerial images for the purpose of updating road maps. Concepts are concretized and tested on a case handling the extraction of new roads linked to existing motorways in large-scale aerial photographs. In the first chapter the demand for automation of photogrammetric processing is discussed as well as complications when using traditional image processing techniques for this task. As a result the need for knowledge-based concepts becomes clear.

1.1 WHY PHOTOGRAMMETRY NEEDS COMPUTER VISION

Developments in photogrammetry have always been closely related with developments in other fields of science and technology, as was pointed out by several authors, [e.g. Schenk 1988, Torlegård 1988]. Progress in cartography and computer science accelerated and increased the interest in the current transition from analytical to digital photogrammetry. Relevant changes in cartography will be discussed first, followed by their consequences for photogrammetry, which will lead to reasons why new techniques in computer science need to be investigated.

In modern map production a shift took place from maps stored in analogue form on paper or film to digital databases containing topographic information. Digital topographic databases are an essential part of Geographic Information Systems (GIS). GIS supports the integration of topographic information with other types of information, like administrative and thematic. Besides it provides a number of sophisticated software tools, for instance for analysis and presentation of spatial data. This makes GIS into a powerful instrument for the purpose of planning, monitoring tasks and management. Moreover it makes new technological developments possible like in-car navigation, where a car is equipped with a small computer which plans and displays the route to a certain destination by using digital maps and other information like locations of traffic-jams. Because of these advantages of GIS compared to paper maps, there is a wider use of topographic information and therefore a larger demand.

However, to be effective, GIS is dependent on accurate and up-to-data input data. This does not only apply for information users add themselves, but also for topographic information. The next examples will make this clear:

- If a traffic accident happens on a newly constructed road, the service that registers these accidents will need up-to-date road databases.
- Drivers who use in-car navigation will demand new roads to be included in their route-planning as soon as possible.

- Public utilities will need newly constructed houses to be present in their topographic database when they deliver and map their services.

These are reasons why users strongly ask a more frequent updating cycle of topographic information.

Photogrammetry established itself during this century as an efficient surveying and mapping method. High-quality and up-to-date topographic information can be extracted from aerial photographs. However, photogrammetric processing forms the bottleneck in speeding up the topographic information supply. Especially measuring three dimensional coordinates by outlining manually every object in the photograph is very labour-intensive and time-consuming. Therefore, further automation of photogrammetric processing is highly desirable.

One approach for automation is to consider every task in the photogrammetric processing chain and try to automate each of them [Heipke, 1993]. An extended overview of the state-of-the-art of automation of photogrammetric tasks is given in appendix A. Summarizing, geometric tasks, such as aerial triangulation and orientation, can at present nearly be solved automatically by transferring experience from analytical to digital photogrammetry. However, tasks involving interpretation capabilities of human operators are very difficult to solve by computers. In particular interpretation tasks are very labour-intensive and time-consuming tasks in the mapping process. In this thesis the notion "interpretation" refers to determination of the location and outlining of objects in the image as well as recognition and classification of topographic objects. Even unexperienced people can immediately recognize for example most roads and houses in aerial images, but nobody can tell exactly how they did it. People unconsciously rely on knowledge about properties of objects and their appearance in the aerial image. In order to perform this task by computer this knowledge should be formulated exactly, together with techniques to measure them in the image, since for a computer a digital image is only an array with numbers, representing grey values.

A current development is to skip the mapping process and to integrate up-to-date image data directly in GIS [Ehlers et al. 1989, Fritsch 1991]. In this way the user immediately possesses new image data. Old GIS-information can be compared with the new situation and if necessary updated by the user himself. As a result of the growing awareness that up-to-date imagery offers good prospects for GIS to be more effective, many commercial GIS products have been adapted to offer image display capabilities and some tools for image analysis [e.g. Laan 1991]. Thanks to these possibilities for integration of GIS and image data another approach for automation of photogrammetric processing became feasible: utilization of information from GIS to improve the automatic extraction of new information from the image data for GIS-updating. Various research shows that ancillary geographic information can improve satellite image classification for thematic mapping, like land cover classification [e.g. Wilkinson/Burril 1991, Janssen 1994]. Also topographic mapping can benefit from GIS information [Cleynenbreugel et al., 1991]. It seems to be a promising approach to solve the very hard task of interpretation by computers.

Interpretation of digital images in general is the subject of computer vision. A definition of this discipline given is by Haralick and Shapiro [1992a]:

Computer vision is the science that develops the theoretical and algorithmic basis by which useful information about the world can be automatically extracted and analysed from an observed image, image set, or image sequence from computations made by special-purpose or general-purpose computers.

It includes techniques from many disciplines which can also be useful by themselves, like digital image processing, statistical pattern recognition, and artificial intelligence. Applications can be found in many areas like medicine, biology, robotics and remote sensing. Photogrammetry can highly benefit from experiences with image interpretation in other fields. For example:
- Image processing techniques suitable for tracking blood vessels in the medical domain can probably also be used for tracking roads.
- Artificial intelligence techniques to order and control processing steps and represent properties of objects will certainly be valuable when building systems for interpretation of aerial images.

Summarizing, the increasing need for up-to-date topographic information in GIS requires faster photogrammetric processing. Since especially the interpretation task is hard to automate, similar experiences in computer vision with this problem can help to find solutions for digital photogrammetry.

1.2 CONTRIBUTION OF COMPUTER VISION TO AERIAL IMAGE INTERPRETATION

Aerial photographs are more complicated for interpretation than for example images taken in an industrial environment, often used in computer vision applications. Their specific characteristics have direct consequences for the suitability of concepts and requirements on strategies from computer vision to solve photogrammetric problems. In this section these complicating factors will be presented together with consequences for computer vision.

1.2.1 COMPLICATING FACTORS FOR INTERPRETATION OF AERIAL IMAGES

The most important reason for disappointing results using traditional digital image processing techniques is the complex contents of aerial images:
- Images of natural scenes contain many objects which occur close to each other or even partly overlap. Only a limited part is of interest for mapping.
- Many of the objects in the aerial image are a complex composition of parts. For example, a road network consists of carriageways, traffic lanes, slip-roads, junctions, fly-overs, etc.
- Objects in general also belong to more specialized classes. For example, a road can be a motorway, main road, street, etc., dependent of its functions.
- Some objects cannot be treated as independent objects, but become meaningful in their context. A bridge for example is recognized as a part of the road which crosses the river.

- Objects belonging to the same class, like houses, can appear in aerial images in a wide range of representations, a different context and on different scale.

Two other sources also contribute to the complexity of aerial images interpretation:
- image acquisition
- computer limitations

Aerial photographs are taken under different conditions: season, weather, time, and altitude. Different seasons and weather cause variations in grey value, colour and texture. Time of image acquisition is related to the sun angle, which causes corresponding shadows. Altitude and focal length of the camera determine the scale of the photograph.

A difficulty that the computer should handle is the fact that scanned aerial photographs occupy a lot of disk space and memory. A photograph of 0.23×0.23 m^2 scanned at a resolution of 100 μm will take about 5 Mb and at a resolution of 10 μm even more than 500 Mb. In addition, there is usually a complete block of photographs of an area and photographs may be taken in different spectral bands. Intermediate results of image processing will also require multiple storage capacity compared to the raw data.

1.2.2 CONSEQUENCES FOR INTERPRETATION OF AERIAL IMAGES BY COMPUTER VISION

Förstner [1991] states that we should not worry about necessary hardware too much as because of the rapid developments in computer science it will be available in speed and storage as soon as we have specified what such systems are to be used for. For the time being it is sensible to apply processing to restricted regions of interest. This is also advantageous when dealing with many objects occurring close to each other in aerial images. Only a limited part is of interest for mapping. By restricting the search area, features of uninteresting objects need not to be considered. Most knowledge-based systems have such facilities.

The bottleneck in the traditional three step paradigm of segmentation, feature extraction and classification (see section 2.1.1) is the first step: segmentation. Separation of topographic objects from their background is very hard due to among others partly occlusion, variations in radiometric properties and texture. Since segmentation and feature extraction followed by classification heavily depend on each other, results will never be optimal for every object. Consequently, a flexible control structure is required which allows combination of several segmentation techniques and if necessary performs re-segmentation after feature extraction or classification [Kestner/Rumpler, 1984]. A control strategy from computer vision should be chosen which fulfils this condition. Selection and combination of segmentation techniques requires knowledge about characteristics of image processing techniques to be included [Matsuyama, 1987].

The wide variety of objects and properties requires the use of knowledge-based concepts from computer vision, including explicit representation of knowledge about objects in the scene. However, building a formal model that includes all relevant knowledge about objects will be a difficult task. Firstly, properties of objects need to be collected from various sources and subsequently be evaluated for their relevance. Secondly, the knowledge representation formalism is required to organize and represent a wide variety of properties, which belong to objects themselves or to the interrelationships between objects.

Map data of the area under investigation is recognized to be quite valuable to locate objects in a complex situation [Matsuyama 1987]. The map information can be considered as a good model for the real situation even though it is out of date. Properties of unchanged objects can be used to search new objects of the same class. Maps could also be used to guide image interpretation based on expected contextual relationships between objects in the map and in the image. Cleynenbreugel et al. [1990] proved that it can be profitable to incorporate knowledge from maps. Nevertheless, maps have hardly been used as knowledge source in automated image interpretation. A consequence of the use of maps is that both a priori general information about objects and information belonging to specific objects in a certain scene need to be included in the interpretation strategy.

1.3 THESIS SCOPE AND CONTRIBUTION

Within the context of automating the extraction of topographic objects from aerial images for map production, this thesis focuses on the potential of knowledge-based concepts for this task. The aim is to design and evaluate an interpretation strategy (fig. 1.1) which is based on a priori knowledge about both topographic objects and image processing techniques together with information from an outdated or incomplete map of the scene under investigation. The result of interpretation can be used to update this map. First, the scope of this thesis and requirements concerning the input data (knowledge base, road map and aerial image) will be discussed.

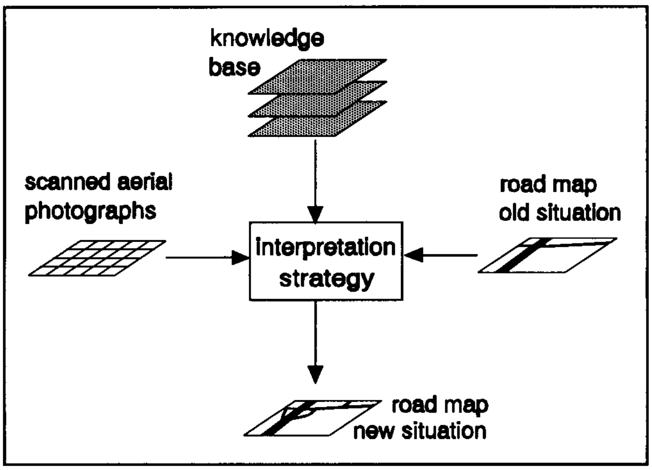

Fig. 1.1 Input and output diagram for knowledge-based interpretation of aerial images for map updating

Especially the contents of the knowledge base will be emphasized. This knowledge base should reflect the complex contents of aerial images. Thus the object model must be able to include, besides properties of topographic objects, also relationships between objects which represent context, specialized classes and component parts. Because of the complexity to build a knowledge base which includes complete descriptions of all topographic objects and their interrelationships, the interpretation strategy is mainly designed for the extraction of one class of topographic object: road networks. Roads are a means of communication between different locations. Since they express the principle structure of an area, they can form a logical framework to search other topographic objects. Even though only roads are considered in this thesis, the design of the knowledge base is required to allow easy extension for other topographic objects. This requirement matches with the design philosophy of most knowledge-based systems. The complexity of interrelationships within this class of topographic objects is comparable to other classes and to relationships between classes. Models for roads are less complicated than for most other topographic objects, because knowledge about the three-dimensional shape, which varies depending on viewing angles, does not play an important part for flat objects like roads. Consequently the use of single images instead of stereo pairs is feasible for extraction of road networks. An original contribution is the insertion of standards for the construction of road networks in the road model and the execution of tests to determine their contribution.

The development of specific low level digital image processing techniques is of minor concern. Choices are mainly based on availability and easy implementation. The presented techniques are fully automatic. However, the design of the interpretation strategy is required to be flexible in order to make it possible to replace them by other segmentation techniques, among which semi-automatic digital techniques. Therefore, choices for image processing techniques and knowledge about optimal parameter settings should be expressed in the knowledge base.

Matsuyama [1987] notes four difficulties when using maps for image understanding:
1. processing of analogue map data in order to extract digital information;
2. establishing accurate image and map correspondence;
3. design of a data structure to store maps during image processing;
4. design of a map guided interpretation strategy.

Because digital road databases are nowadays often available and their availability will grow in the future, the first problem is not within the scope of this thesis. Correspondence between our road databases and aerial images can be established by using manual indicated points and orientation parameters calculated during aerial triangulation. The correspondence problem will not be discussed further in this thesis and also automation of this task is not within the scope of this thesis. The third and fourth problem will be highlighted within this thesis and yield another original contribution. The interpretation strategy is based on the assumption that the road database is outdated or incomplete and uses knowledge about possible changes.

In many previous work on road extraction described in literature satellite images or small-scale aerial photographs are used. At this scale the road model is rather simple: a network of lines and intersections. A requirement for the road model and interpretation strategy in this thesis is that they have the potential for use at several scales. Especially the use of large and medium scale aerial photographs and the objects that need to be mapped at those scales, will yield an original contribution.

The Survey Department of the Ministry of public works, water management and transport in the Netherlands is interested in road extraction since one of their tasks is to keep a digital database of the Dutch motorways up-to-date. Their users demand more frequent updating. The case study in this thesis, which is used to concretize and evaluate the concepts, concentrates on updating of road databases using scanned aerial photographs from their practice. Results of tests on this data will be presented.

1.4 THESIS ORGANIZATION

The rest of this thesis is divided into three parts. The first part, chapter 2, 3, and 4, deals with existing approaches towards image understanding in general and road extraction in particular and discusses the designed knowledge-based interpretation strategy at a conceptual level. Because concepts from computer vision like knowledge representation and reasoning strategies will play an important role, chapter 2 discusses the theory of these issues. Readers familiar with these topics may skip this chapter. Chapter 3 gives a review of previous work on road extraction. The interpretation strategy developed in this thesis for updating of topographic objects in general and road extraction in particular will be outlined in chapter 4.

In the second part, chapter 5 and 6, the concepts are concretized by a case study that matches with the photogrammetric practice at the Survey Department. Large-scale up-to-date photographs (scale 1:4000) of parts of the Dutch motorways are used together with outdated digital topographic databases. One situation is regarded: the extraction of new roads linked to existing motorways. Chapter 5 describes the different road models that are used in the case study. Results of tests with these models are presented and discussed in chapter 6.

The last part, chapter 7, gives conclusions about the suitability of the developed interpretation strategy for updating of road maps and discusses the potential of knowledge-based road extraction for the photogrammetric practice. Finally, recommendations for further research are given.

PART I

THEORY AND CONCEPTS

In this part we discuss approaches for extraction of roads from aerial images that were described in literature and present concepts of the designed interpretation strategy.
A conclusion from previous work is that a priori knowledge about the objects in the scene and their context are required. Relevant issues from artificial intelligence, like knowledge representation and reasoning strategy, are evaluated and selected for the purpose of aerial image interpretation. They are incorporated into the strategy we develop for updating of road maps from aerial images.

CHAPTER 2

CONCEPTS IN KNOWLEDGE-BASED IMAGE INTERPRETATION

In the first chapter reasons were given why computer vision can contribute to automation of photogrammetric processing. Computer vision is confronted with similar problems when trying to automate the interpretation task. It seems reasonable to analyse its concepts, strategies and algorithms with the aim to judge which concepts may be useful for solving photogrammetric problems. Therefore, we need to know what approaches have proven to be successful and under which conditions. An overview of computer vision, its paradigms, basic concepts and strategies is given in this chapter, meant for readers not familiar with those topics.
Because knowledge-based techniques are most promising for interpretation of aerial images, these techniques will be used within this thesis. This chapter will give an overview of their most important concepts and approaches. Several strategies to control the interpretation process and to represent a priori knowledge will be discussed.
General text books about computer vision, like [Ballard/Brown, 1982], [Gonzalez/Woods, 1992], [Haralick/Shapiro, 1992a+b] and [Sonka et al., 1993] and books on artificial intelligence, like [Barr/Feigenbaum, 1981] and [Rich/Knight, 1991] provide more details about these subjects. Binford [1982] provides an extensive survey of several knowledge-based systems.

2.1 IMAGE INTERPRETATION BY COMPUTER VISION

2.1.1 TRADITIONAL STRATEGIES FOR IMAGE INTERPRETATION

Difficulties in computer vision mainly arise from the lack of fundamental processing tools to get from what is given (an array of pixels) to what is desired (a symbolic representation of the image content). The first approaches to analyse images come from the field of digital image processing and pattern recognition.

A commonly used approach to analyse images is the straightforward three step paradigm: segmentation, feature extraction, classification (fig. 2.1). For each of these steps there is a large assortment of digital image processing techniques to choose from. Most of them are domain independent. Examples are techniques like edge detection and thresholding for segmentation and using features like area, length or curvature for classification. It is often a matter of trial and error to find for a certain task a feasible sequence of techniques and their parameter settings. Within their limited task domains and in a controlled environment, like homogeneous background and same illumination, the three step paradigm has proven to be successful [Groen/Munster, 1986]. However, if the environment changes, image processing has to be adapted as well. The condition of a controlled environment is not fulfilled for aerial images, for example because the same type of objects may appear on various backgrounds. Results of this approach on aerial images [e.g Bajcsy/-Tavakoli 1976, Wang/Howarth 1990] confirm that it is not feasible for map making.

An inconsistency of the traditional three step paradigm is that general algorithms and features are used, which use very little knowledge about the domain, whereas special purpose systems are build. As a consequence it is often not clear which parameters of the algorithm correspond to which properties of an object and actually lead to recognition. This fact contributes to the exhaustive trial-and-error process necessary to develop a method. Often conditions like a black object on a white background are defined implicitly within the algorithm. If the environment changes, this may require a fundamental adaptation of image processing techniques. This resulted in the design philosophy to represent domain-dependent and object-specific knowledge explicitly and separate this knowledge from general problem-solving computation [Hanson/Riseman 1978, Draper et al. 1989]. This philosophy provides maximum flexibility during development of the system and permits modification for other applications. This has led to the notion "knowledge-based" systems, also referred to by the notion "model-based" systems.

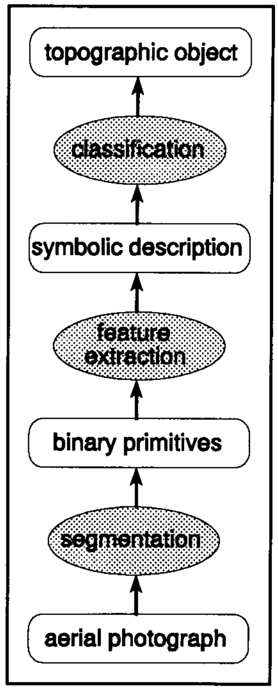

Fig. 2.1 Traditional three step paradigm for image analysis

2.1.2 KNOWLEDGE-BASED IMAGE INTERPRETATION

A priori knowledge is organized to form models or knowledge bases for each object class. These models should clearly show the involved object properties. It has been proven that it is very important to incorporate also more advanced features describing the context e.g. spatial relationships between objects and object-parts. The choice of a knowledge representation formalism can seriously affect the performance of the system. Possibilities will be reviewed in section 2.5.

Only if the domain is extremely simple and heavily constrained, the object knowledge can be matched directly with the image (e.g. using template matching). In other cases digital image processing techniques should extract information to fill in a symbolic description of the object, which can be matched with the object knowledge. It may be necessary to compose new structures by grouping, splitting and/or modification.

More complicated problems, like interpretation of natural scenes, require a combination of several digital image processing techniques. One reason is that the scene often contains a large number of different kind of objects with their own features [Wong/Frei 1992]. Another reason, with drastic consequences, is that there is no set of parameter settings for any algorithm that will extract the desired information perfectly [Matsuyama 1987]. Alternative techniques should be provided to cope with errors in a flexible way. As a consequence the sequence of processing steps should allow iterative refinement of results, adaptation of parameters and feedback. All these conditions require a strategy that controls application of knowledge bases, activation of image processing techniques and orders processing steps. Control strategies are a topic of artificial intelligence and the suitability of several alternatives for analysis of aerial photographs will be analysed in more detail in section 2.3.

2.2 LEVELS OF PROCESSING AND REPRESENTATION

Computer vision has to deal with a wide range of processes and representations to derive a meaningful description of the scene from an array of pixels. In order to structure these processes and representations the terms low level and high level were introduced. One should notice that these terms are used with different meanings: as levels of processing or as levels of representation. Fig. 2.2 shows both hierarchies with their data representations and processing tasks.

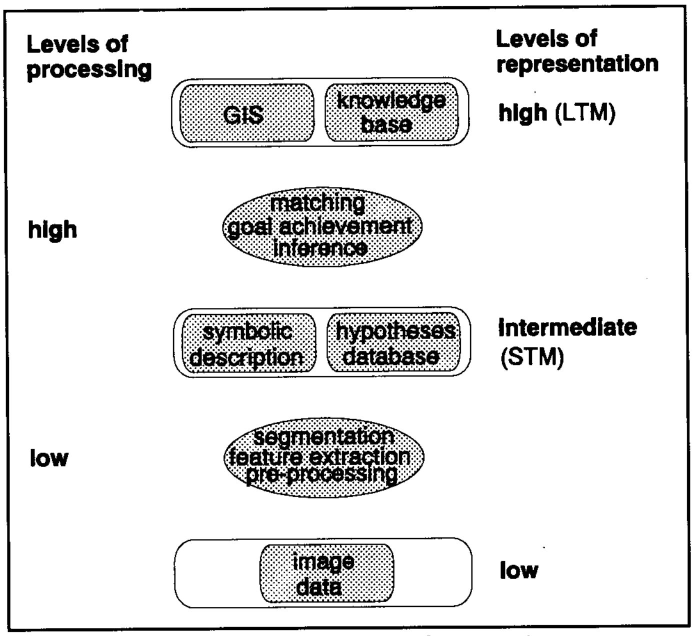

Fig. 2.2 *Levels of representation and processing*

Ballard and Brown [1982] and Sonka et al. [1993] distinguish two levels of processing: low and high. Low level processing operates on digital image data and uses very little knowledge about the contents of the image. It performs tasks like pre-processing, segmentation, and feature extraction. High level processing is based on knowledge about the application and goals for interpretation. It includes capabilities like making plans how to achieve those goals, ordering of low level image processing steps, matching of formalized models of the world with the image content and inference of sub-goals for further processing. Which processing will be performed strongly depends on the selected control strategy and will be discussed in the next section. Other division of techniques into levels of processing are used as well, e.g. into three levels by Gonzalez and Woods [1992], but the division into two levels of processing fits best with the hereafter defined levels of representation.

Hanson and Riseman [1988] and Förstner [1993] define a hierarchy of three levels of representation: a low, intermediate and high level. At the low level image data is stored. This can either be one image or multiple images e.g. of different resolutions, spectral bands or view points. At the high level domain-dependent a priori knowledge and models are stored. Which types of knowledge can be distinguished and alternatives for knowledge representation will be the subject of section 2.4 and 2.5 respectively. If in addition a digital map is used, it should be stored at the high level as well. The reason is that this map is not only the final result of interpretation, but information from outdated digital maps can be considered as a priori knowledge. The high level of representation is often called the long-term memory (LTM), because it embodies knowledge which needs to be stored for a long time and can be used for several scenes. An additional intermediate level is defined to store for a specific scene its symbolic representations for regions, lines and surfaces with their features that have been extracted from the low level image data. Since it contains instantiations of objects in a specific scene, which only need to be stored temporarily, this level is sometimes called short-term memory (STM). The set of hypotheses constructed from the knowledge at the high level applies also for a specific scene, so it should be stored at the STM as well.

2.3 CONTROL STRATEGIES

The choice of a control strategy dictates the direction of the flow of information between the different levels of representation. Definition of a particular representation level as input or output level results in quite a different type of processing and consequently influences fundamentally the interpretation process.
There are two major approaches of control [Haralick/Shapiro, 1992b] which are discussed in more detail: hierarchical control and non-hierarchical, also called heterarchical control.

2.3.1 HIERARCHICAL CONTROL

In this context hierarchy refers to levels of representation. In hierarchical control two extremes can be distinguished:

- Bottom-up control or image data driven: from the low level to the high level of representation
- Top-down control or model driven: from the high level to the low level of representation

Bottom-up control overlaps with traditional image processing techniques. First, segmentation of the image produces binary primitives, such as lines or regions. Next, by feature extraction a symbolic description of these primitives is constructed. Finally, classification based on these features leads to recognition of topographic objects. Figure 2.1 already showed the general outline of bottom-up processing. Input and output data are represented by boxes and processing tasks are represented by ellipses.

There is no standard version of top-down control as presented for bottom-up control. A general top-down process is visualized in figure 2.3. The general mechanism of top-down control is hypotheses generation and its testing. Processing always starts with generation of a set of hypotheses, based on stored knowledge about the object to be recognized. An object hypothesis is defined as a statement about the presence of an object and can be either true or false. The task to determine the sequence of object hypotheses, in order to arrive at (or infer) an interpretation of the scene, is called inference. The next step, usually, is to use for each object hypothesis a focus-of-attention mechanism, which constrains for example the part of the image to be processed and the range of attributes of extracted primitives. This information is input for a goal-directed segmentation. If there is any accepted output, it is supposed to be of the hypothesized type of object.

Vision systems based on pure top-down control do not exist. A common approach is to use currently extracted objects in a bottom-up fashion to generate new hypotheses and to adjust the sequence of the set of hypotheses.

Hybrid control mechanisms, that combine both bottom-up and top-down control strategies, usually give better results than either basic control strategy applied separately. Fig. 2.4 gives an overall scheme which integrates both bottom-up and top-down control. A common hybrid strategy is to start with a bottom-up initial segmentation of the image and extraction of a preliminary set of features and relationships. On the basis of this preliminary symbolic description, the identity of one or more objects is hypothesized. Now a top-down strategy can be used to verify or disprove the existence of these objects.

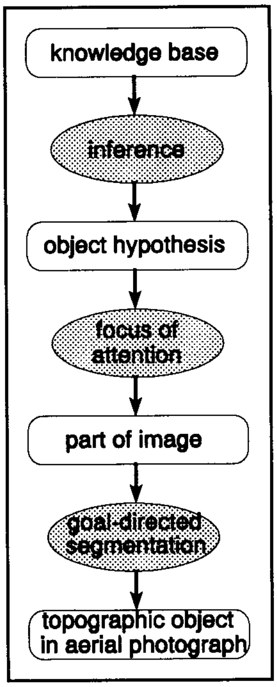

Fig. 2.3 Top-down control

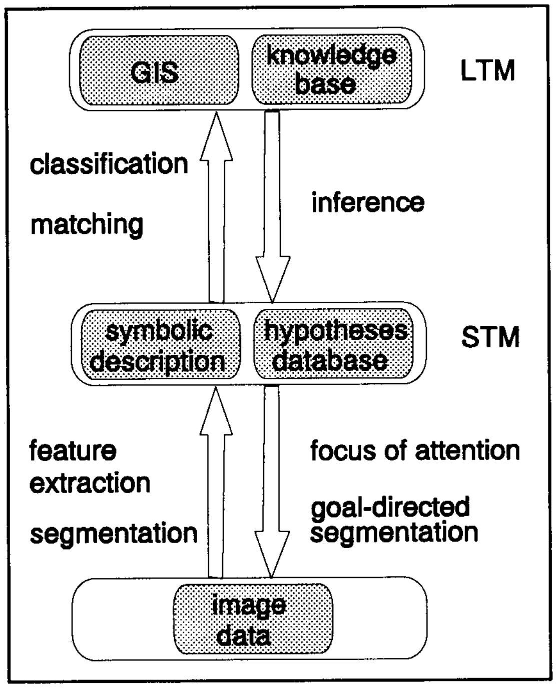

Fig. 2.4 Hybrid control, a combination of bottom-up and top-down

Knowledge about relations of these verified objects allows new hypotheses to be generated. For example, if parts of objects are recognized, more information can be deduced to build the complete object. Or knowledge about spatial relationships in the scene can predict location and class of other objects in the neighbourhood of the recognized object.

2.3.2 HETERARCHICAL CONTROL

Rather than looking at the levels of representation as a hierarchy, the current state of the data and acquired information can be seen as activator of knowledge sources operating at the same level, called heterarchical control. Knowledge sources are independently executable procedures that contain domain-specific knowledge. Each knowledge source can communicate with some or all of the other knowledge sources. For example, the knowledge source to detect shadows can activate the knowledge source to extract houses, which in turn can activate the knowledge source to track roads. Because the objects present in the scene dictate activation of knowledge sources, the order in which the expertise should be deployed is not fixed. Hence it is difficult to keep track of the interpretation process.

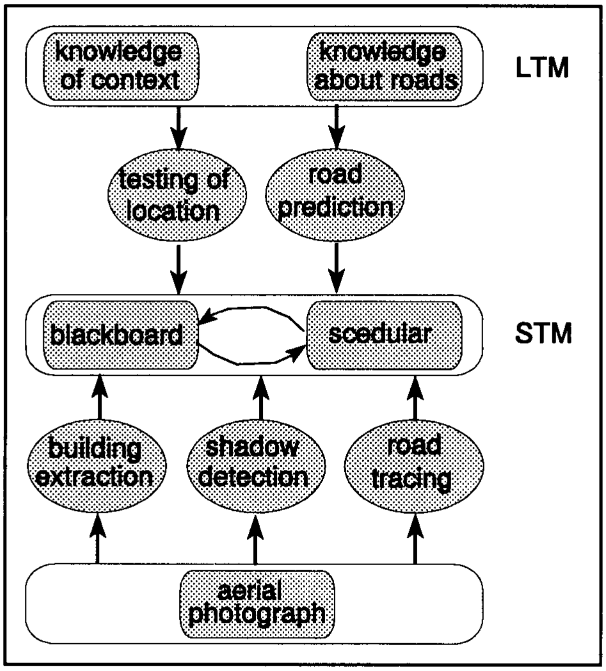

Fig. 2.5 Heterarchical control by a blackboard model

With the purpose to add some order in the heterarchy, the blackboard approach was introduced. Figure 2.5 shows an example of a part of a possible structure of a blackboard for aerial image interpretation. All communication between knowledge sources has to take place via a shared database, called blackboard, which stores all the information extracted by knowledge sources. The basic idea of the blackboard can be explained best by imagining a classroom full of cooperating and competing experts, called knowledge sources. Each expert in turn can try to contribute to interpretation of the scene if its preconditions, associated with each knowledge source, are met. A kind of schoolmaster, called blackboard schedular, determines the order of execution of competing experts. He asks assistance of the expert that can probably help most to obtain the final solution. An expert can extract primitives or features from the image or he can generate hypotheses or verify information using the knowledge base. Results are written on the blackboard and can be used by other experts in this way working towards an incrementally developing interpretation.

2.4 TYPES OF KNOWLEDGE

A knowledge-based system is organized in such a way that knowledge about the problem domain is separated from general processing routines. This collection of domain-dependent knowledge is called the knowledge base. A general property of this knowledge is that it needs to be specified in advance and needs to be stored for use in multiple scenes. Hence it is called a priori knowledge and is represented in the LTM.
Properties of knowledge which are not generally valid need to be considered as well when choosing an appropriate representation to formalize the knowledge. Therefore, different types of knowledge will be defined in this section before alternatives for representation are discussed in section 2.5.

The main classification is based on the nature of the knowledge. Distinction is made between declarative and procedural knowledge.

Declarative knowledge specifies what is known about the task or about the objects to be recognized. Examples of declarative knowledge are:
- A motorway has more than two lanes at each carriageway.
- A lane is about 3,5 meters wide.

Procedural knowledge specifies how to perform a task, for example:
- To calculate the width of a road, divide its area by its length.
- To find a bridge, search at positions where roads cross rivers

Both types of knowledge will be discussed in more detail in order to emphasize other properties within these classes that influence the representation. Fig. 2.6 shows an overall scheme of the discriminated types of knowledge.

2.4.1 DECLARATIVE KNOWLEDGE

There are many possibilities to subdivide declarative knowledge. In this section types of knowledge are discriminated if they influence the choice of a suitable representation. For this reason a distinction is made between so-called object knowledge and relational knowledge. The first describes knowledge about properties of individual objects, the second relational constraints among objects.

Sometimes it is profitable to define classes of objects which contain general object knowledge. Specific classes, sometimes called "children", inherit properties from general classes, called "parents". Children add their own discriminating properties. In order to support inheritance, objects must be arranged into classes and classes must be arranged into a generalization hierarchy. Object knowledge arranged in such a hierarchy will be called inheritable knowledge [Rich/Knight, 1991], in contrast to non-inheritable object knowledge, which does not require organization of objects into classes.

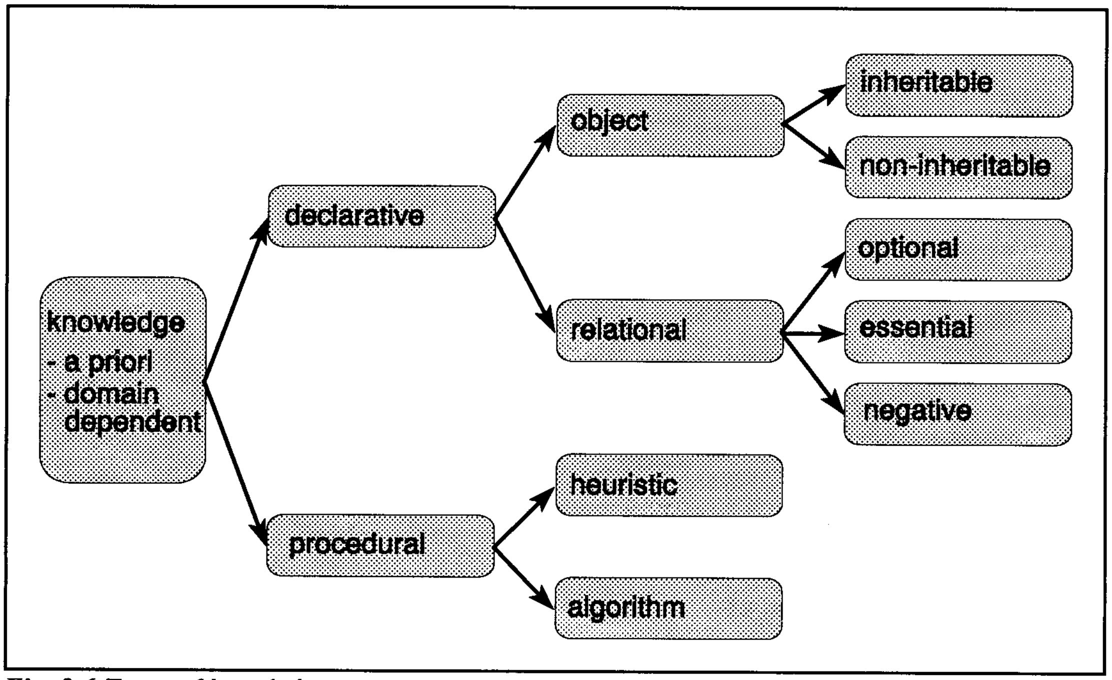

Fig. 2.6 Types of knowledge

In [Hartog, 1995] three types of relationships are distinguished:
- optional;
- essential;
- negative.

Optional relationships are relations between two objects that are likely to be present, but not necessarily. A factor indicating this likelihood may be attached, which can be used as a control mechanism. Essential and negative relationships both define constraints among objects which respectively either essentially need to be present or are not allowed to be present. They can be used to detect inconsistencies.

2.4.2 PROCEDURAL KNOWLEDGE

The procedural knowledge which should be represented explicitly is domain-dependent. Hence inferential knowledge in order to exploit declarative knowledge will not be considered. Procedural knowledge deals with operations like selecting image processing routines and setting parameters for these routines. It consists of both heuristics and algorithms.

Algorithms produce the correct or optimal solution to a problem, based on physical laws. For example, to model aspects of image formation, such as the projective transformation within a camera, algorithms are used with domain-dependent parameters for among others focal length and flight height.

Heuristics are based on experience and aim to limit the search for solutions, however, often there is no guarantee that they produce the correct solution. Image processing routines like line detectors are examples of heuristics for the recognition of roads since not every line will be a road.

2.5 TECHNIQUES FOR KNOWLEDGE REPRESENTATION

A knowledge-based approach raises the problem of choosing a formalism to express knowledge. Requirements for the formalism used in the knowledge-base are:
- it should be flexible in integrating new knowledge into the existing knowledge-base;
- it should be shown in a form which is easy to read;
- it should encourage to separate domain-dependent knowledge from general processing knowledge;
- it should be able to cope with the types of knowledge discriminated in section 2.4.

The most common formalisms to represent knowledge in systems for computer vision are:
- production rules
- semantic networks
- frames or schemas

For each of these types their syntactic and semantic conventions will be described and further illustrated by expressing knowledge with this formalism for the following example:

> *Using the represented knowledge the aim is to recognize which extracted roads can be classified as motorway. Suppose the only criterium for a motorway is that it is a road with more than two lanes. If lanes are not yet defined as parts of the road, a procedure needs to be activated which extracts regions. Suppose it needs only a criterium for homogeneity, defined as a grey value variance of 5.0. The extracted homogeneous regions are lanes if they are elongated and about 3.5 meter wide. Finally, the number of lanes being part of each road needs to be counted in order to determine which roads are motorways.*

Since knowledge representation affects data and processing in the highest levels of representation, the contents of LTM and STM and high level processing will be described based on example systems in which this representation type is used. Which of the above mentioned requirements are met for each of the formalisms, will be discussed at the end of this section.

2.5.1 PRODUCTION RULES

Syntax and semantics

Production rules are expressed as condition-action pairs and have a standard form:

IF <conditions> THEN <actions>

where the conditions and actions can be expressed as conjunctive clauses.
Fig. 2.7 represents the knowledge in the example as production rules.
This form clearly shows the procedural character of the knowledge represented by production rules. Production rules about which paths are most likely to lead quickly to a goal state can be used as a control mechanism. A standard method is to add certainty factors to the rules expressing probability of success. In this way the difference between heuristic and algorithmic knowledge can be characterized.

Rule #1
  IF a road consists of more than 2 lanes
  THEN the road is a "motorway"

Rule #2
  IF a homogeneous region:
    - is about 3.5 meters wide AND
    - is elongated
  THEN classify the region as "lane"

Rule #3
  IF the length of a region is at least 10 times larger than its width
  THEN a region is "elongated"

Rule #4
  IF homogeneous regions are not yet extracted
  THEN activate the procedure "Region_Extraction(homogeneity)" with
    a variance of 5.0 as criterium for homogeneity

Fig. 2.7 Knowledge for classification of highways represented as production rules

Example systems
A knowledge-based system using rules is called a production system. Production rules are the most popular type of knowledge representation technique in expert systems [Waterman, 1986]. Nazif and Levine [1984] describe an expert system for low level image segmentation guided by rules. Rules were used as well to represent specialized procedures to locate specific objects in aerial images [Nagao/Matsuyama, 1980]. SPAM is a rule-based system to interpret aerial images of airport scenes [McKeown et al., 1985].

Representation in the LTM
Since production systems generally contain many rules, knowledge represented in the LTM is often organized in classes. Nazif and Levine [1984] for example discriminate three different types of rules:
1. Knowledge rules, describing object properties.
2. Control rules, which can be divided into:
   a) Focus-of-attention rules, defining the sequence in which STM primitives will be checked.

b) Meta-rules, defining the order in which different knowledge rules will be matched.
3. Strategy rules, selecting a set of control rules.

Representation in the STM
There is no specific form in which processing results of actions defined within the rules are stored in the STM. Production systems often use a blackboard as STM [Gonzalez/Woods, 1992]. In this case the knowledge sources are rules. Nagao and Matsuyama [1980] use tables with properties and parameters to represent STM data on the blackboard. Nazif and Levine [1984] store three types of primitives together with their features in the STM: regions, lines and areas, which are aggregates of regions and lines with certain properties. McKeown et al. [1985] even define a hierarchy of primitives in the STM, in which a primitive in the current level is an aggregate of primitives of the lower level. From the lowest up to the highest level these primitive are called regions, fragments, functional areas and models.

High level processing
Two different ways in which rules can be used in a production system are forward chaining and backward chaining, corresponding to bottom-up and top-down reasoning respectively.
A production system based on forward chaining matches rules in the LTM against the symbolic data stored in the STM. When a match occurs, the rule fires. This triggers an action to be executed which usually involves modification of data in the STM. If data in the STM is changed, conditions of rules in the LTM need to be matched again to check whether other matches occur. If more than one match occurs, one rule is selected using control rules.
When using backward chaining a set of rules that leads to a (sub)goal is selected. Other rules are searched of which the action parts yield conditions required by the first set of rules. This process is repeated until all conditions are fulfilled by the current state of the data. Then the complete processing chain leading to the (sub)goal is executed. If alternative processing chains are formed, certainty factors can help to select one chain.

2.5.2 SEMANTIC NETWORKS

Syntax and semantics
Semantic networks were first introduced under that name as means of modelling human associative memory [Quillian 1968], but are now a standard representation method in computer vision. A semantic network represents objects and relations between objects as a graph structure, i.e. a set of nodes connected by labelled arcs. Nodes usually represent objects and arcs represent relationships between nodes. Common arcs are "is-a" and "has-part" relations. The first one establishes inheritance in the network. Semantic networks describe knowledge in a declarative fashion.

The a priori knowledge in the example that a motorway is a road which consists of at least two lanes of about 3,5 meter wide can be modelled by the semantic network in fig. 2.8.

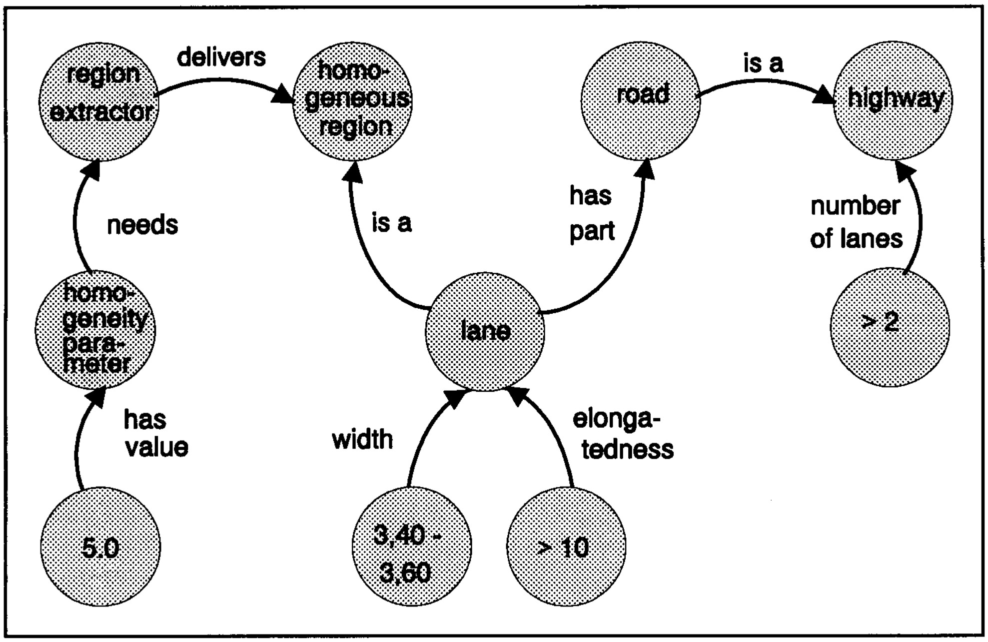

Fig. 2.8 Knowledge for classification of highways represented as semantic network

Example systems

Semantic networks were already early used in aerial image applications by Nevatia and Price [1982] to construct a map-like sketch of the area which guides the segmentation. Nicolin and Gabler [1987] apply semantic networks to represent knowledge about prototypes of scenes and processing methods for the detection of houses and roads in aerial images of suburbs.

Representation in the LTM

Nicolin and Gabler [1987] subdivide knowledge in the LTM in two partitions, both represented by a semantic network:

1. meta-knowledge about processing methods;
2. generic knowledge about suburban scenes.

The first partition includes knowledge about existence, purpose, and interfacing. Processing methods, method-specific parameters, and the kind of input/output data are represented as nodes in the semantic network.

The second partition represents several different types of objects and scenes, structured by two hierarchies of relations:

1. specialization and generalization relations, called "is-a" and "is-generalization-of" relations respectively, which provide an inheritance mechanism;
2. composition and decomposition relations, called "is-part-of" and "has-parts" relations, reflecting the construction of complex structures from simple objects.

Nevatia and Price [1982] define in addition to these relations also spatial relationships, which describe proximity (e.g. adjacent, nearby) and relative position (e.g. above, left, parallel).

Representation in the STM
The standard solution to incorporate STM data in the semantic network is to define a specific instance of a generic concept as node and connect it by an "element-of" relation [Ballard/Brown, 1982].

High level processing
Nevatia and Price [1982] as well as Nicolin and Gabler [1987] both apply bottom-up processing and use the semantic network only for interpretation. In the first phase of interpretation, data in the STM can easily be matched with the a priori knowledge in the LTM, because both are represented by a semantic network. In the second phase of interpretation predictions are made and tested for missing objects, composite structures, and spatially related objects. Ambiguities and inconsistencies are reduced by relaxation labelling.

2.5.3 FRAMES AND SCHEMAS

Syntax and semantics
Frames can be considered as a means to assign more structure to nodes as well as to arcs of a semantic network. Thus the complete collection of all defined frames can be represented by the underlying semantic network. In a frame-based system the objects at each node in the network is defined by a collection of attributes, called "slots", and values of those attributes, called "fillers". Slots are either properties of objects or relations, like "is-a" and "has-part", of which the first one establishes inheritance like in a semantic network. Each slot can have procedures attached to it, which are executed when the information in the slot is changed. In this way the consistency within the whole data structure is maintained. Procedures which are often attached to slots are: "if-added", "if-removed", and "if-needed" [Waterman, 1986].
Frames contain both a declarative part and a procedural part and therefore frames occupy a continuum from totally declarative to totally procedural. The term schema refers to "frames" used in a visual context, however, in the literature [e.g. Haralick/Shapiro 1992b, Kasturi 1992] both terms are used interchangeably. The example represented by frames is shown in fig. 2.9.

Example systems
MISSEE [Glicksman, 1983] uses schemas which both control the interpretation process and build the resulting interpretation. In SIGMA [Hwang et al., 1986] frames are used to generate hypotheses about houses and roads in aerial images. The ACRONYM system [Brooks, 1983] uses frames to model 3-dimensional airplanes collecting parameters of generalized cylinders in the frame.

Representation in the LTM
The slots of frames in the LTM contain constraints on the values of properties or relations of objects.

```
frame MOTORWAY
    slots:
        is-a: ROAD
        number-of-lanes: > 2
    procedures:
        if number-of-lanes is needed then count the number of LANES of parent
        ROAD
end-frame

frame ROAD
    slots:
        has-part: LANE
        is-generalization-of: MOTORWAY
    procedures:
        if ROAD has no LANES as parts then search HOMOGENEOUS REGION
end-frame

frame LANE
    slots:
        is_a: HOMOGENEOUS REGION
        is-part-of: ROAD
        width: 3.40 - 3.60 meter
        elongatedness: > 10
    procedures:
        if elongatedness is needed then calculate length(LANE)/width(LANE)
end-frame

frame HOMOGENEOUS REGION
    slots:
        is-generalization-of: LANE
        homogeneity criterion: 5.0
    procedures:
        if HOMOGENEOUS REGION is needed then activate procedure
        RegionExtraction(homogeneity criterion)
        if HOMOGENEOUS REGION is added then investigate if it is a gener-
        alization of a LANE
end-frame
```

Fig. 2.9 *Knowledge for classification of highways represented by frames*

Representation in the STM

Usually instances of prototype objects are stored in the STM by making a copy of the frame and assigning to the slots values which express properties of specific objects.

High level processing

Two different approaches can be discriminated:

1. Segmentation of all image structures and then classification by matching their appearan-

ces with specified attributes, followed by a check of their relations. This corresponds to a bottom-up approach.
2. Use of relationships to predict locations of objects related to already classified objects. This corresponds to a top-down approach.

In SIGMA [Hwang et al., 1986] both approaches are integrated. Processing starts with an initial segmentation to extract image structures by sequentially selecting hypotheses about primitive objects. These hypotheses are constructed from frames. For each primitive which satisfies a hypothesis, an instance is created by making a copy of the frame, of which the hypothesis is constructed. The instance is inserted in the STM. The next purpose is to search new objects related to instances, group instances, or find out to which more specialized class they belong. Procedures defined in the frames of all instances are evaluated and if the condition part is fulfilled, the corresponding action is put into an action list. These actions are scheduled and then subsequently fired. The result is either an instance or nothing.

2.5.4 DISCUSSION

The main difference between the three types of representation just outlined is the nature of the knowledge they represent. Rules are better suitable to represent procedural knowledge, semantic networks on the other hand clearly represent declarative knowledge. Frames or schemas represent both declarative and procedural knowledge and can exploit the strength of both forms. As a result, attributes of objects are easier to read and to modify in a semantic network, but it may be hard to find out to which chain of procedures will be formed for a specific scene. It is easier to find out which procedure will be performed in a production system, but different attributes of the same object are not always stored together. Frames have the advantage that procedures are stored with all attributes of an object class.

Even though it is possible to define a generalization hierarchy of objects in a production system [e.g. McKeown et al., 1985], it is more natural and more clear to use an object-oriented representation, like semantic networks or frames, when inheritable object knowledge is involved. In semantic networks the distinction between object knowledge, represented by nodes, and relational knowledge, represented by arcs, is also clearer. However, representation of essential and negative relationships is clearer in frames, because procedures related to the presence or absence of these relations can be stored together with the conditions.

Because production systems usually contain many rules, modification and incorporation of knowledge is more difficult then for the other formalisms. Organization of knowledge in different modules is essential.

Sometimes several formalisms to represent knowledge are used within one system. Fig. 2.9 already showed how rules can be incorporated in frames. In probably the best known knowledge-based system, VISIONS, described in many publications [e.g. Hanson/Riseman 1978, Draper et al. 1989, Hanson/Riseman 1988] even all three described formalisms are used.

VISIONS operates on natural outdoor scenes. Declarative knowledge is represented in semantic network style. The term "schema" is used to indicated the highest level of object structures in the network. Rules and frames are supported within the VISIONS environment to control processing.

The next chapter reviews previous work on road extraction. In some of it the described formalisms will be used. Both reviews, in this chapter and the next one, will be used to make a final choice for a knowledge representation especially suitable for road network updating. Although the choice of a suitable knowledge representation is very important, the quality of the domain-dependent knowledge still determines the success of a system. In the next chapter also a review will be given of knowledge used to model roads.

CHAPTER 3

REVIEW OF PREVIOUS WORK ON ROAD EXTRACTION

An expansive literature has grown since the beginning of this decade on the problem of interpretation of aerial images in general and extraction of roads in particular. In this chapter a large number of relevant publications on road extraction (21) will be reviewed. If more papers about the same road extraction procedure were published by the same group, for the overview given in this chapter the most complete or latest one was selected. First, an overview of those characteristics is given which influence the complexity of the problem and the suitability and possibility to apply certain image processing techniques. Next, previous work will be reviewed, categorized by the used control strategy. Finally an overview of road properties is made and it will be shown how they are included in previous work.

3.1 OVERVIEW OF CHARACTERISTICS

When reviewing previous work one should realize that successful interpretation not only depends on the strategy and techniques used for road extraction, but also on the type of images on which they are applied. The following two factors affect the complexity of road extraction:

1. road appearance, which depends on factors like ground resolution, contrast with the surroundings and amount of occluded parts;
2. road context, since depending on the scene under consideration (e.g. urban or rural), the road network can have a different degree of complexity in e.g. density and shape (straight, serpentine).

In section 3.1.1 and 3.1.2 overviews will be given of the geometric resolution and type of landscape for which the reviewed publications were developed. Since we are especially interested in knowledge-based approaches, section 3.1.3 lists which publications use this approach. This information will be useful as look up tables to understand some choices made in publications reviewed in the next sections.

3.1.1 ROAD APPEARANCE

Roads have different appearances at varying scales. In satellite images, which are of very low scale, road extraction is usually viewed as linear feature detection. In large scale aerial photographs the details on the road surface are clearly visible, so that a linear element can be decomposed into detectable primitives such as lanes, fly-overs, or crossings. Consequently the scale of the aerial image determines which road model and image processing techniques are appropriate to use.

LANDSAT		SPOT	
author(s)	geometric resolution (in m²)	author(s)	geometric resolution (in m²)
Bajcsy and Tavakoli [1976]	57×79	Cleynenbreugel et al. [1990]	10×10, 20×20
Ton et al. [1989]	30×30	Grün and Li [1994]	10×10
Sijmons [1987]	30×30	Gunst et al. [1991]	10×10
		Maillard and Cavayas [1989]	10×10
		Wang et al. [1992]	10×10

Table 3.1 *Ground resolution of reviewed literature using satellite imagery*

> 10x10 m²		10x10 - 1.0x1.0 m²		< 1.0x1.0 m²	
author(s)	geometric resolution (in m²)	author(s)	geometric resolution (in m²)	author(s)	geometric resolution (in m²)
Fischler et al. [1981]	?	Groch [1982]	1.0×1.0 to 5.0×5.0 *1	Airault et al. [1994]	0.5×0.5 to 1.0×1.0
		McKeown and Denlinger [1988]	1.0×1.0 or 3.5×3.5	Fua and Leclerc [1990]	?
		Vosselman and Knecht [1995]	1.6×1.6	Garnesson et al. [1990]	0.4×0.4 or 0.85×0.85 *2
		Zhu and Yeh [1986]	3.0×3.0 to 4.0×4.0	Heipke et al. [1994]	0.2×0.2 *1
				Hwang et al. [1986]	0.75×0.75 *3
				Lemmens et al. [1988]	0.28×0.28
				Nagao and Matsuyama [1980]	0.5×0.5
				Stilla and Hajdu [1994]	0.16×0.16 *1

Table 3.2 *Ground resolution of reviewed literature using aerial photographs*
**1 = calculated from scale photograph and scan resolution*
**2 = deduced from other paper using the same test images*
**3 = calculated from estimated distances or sizes in reality*

In table 3.1 and 3.2 the reviewed literature on road extraction is categorized by the scale of the aerial images used, indicated by the geometric resolution of test images used. This is the size of the ground area to which one pixel in the image corresponds. Satellite images are subdivided into Landsat and SPOT (table 3.1). Corresponding resolutions depend on the bands used.

The variety in scales of scanned aerial photographs is larger. They are subdivided (table 3.2) into road extraction applied on test images with a ground resolution:
- smaller than or equal to that of satellite images, i.e. larger than 10 metre;
- between 10 and 1 metre;
- smaller than 1 metre.

The ground resolution is not always given directly, but can be deduced or else it is visually estimated to which of the three categories the test images belong.

3.1.2 ROAD CONTEXT

The road context influences road properties and therefore also which image processing techniques are suitable. Table 3.3 lists which types of landscape the test images used in each reviewed publication depict. A discrimination is made between urban, suburban, rural and uncultivated scenes.

3.1.3 KNOWLEDGE-BASED ROAD EXTRACTION

Most of these publications use a traditional image processing approach. Only six of them use a knowledge-based approach. Two types of knowledge representations are used: production rules and frames. In [Nagao/Matsuyama, 1980], [Zhu/Yeh, 1986] and [Stilla/Hajdu, 1994] production rules are used. Garnesson et al. [1990] use their own object oriented language, based on LISP, coupled with production rules, which produces definitions looking like frames. SIGMA, described in [Hwang et al., 1986], is used in section 2.5.3 as an example of a system based on frames. Cleynenbreugel et al. [1990] use an object-oriented environment for image understanding (see [Fierens et al., 1991]), implemented on top of an existing knowledge engineering tool KEE, which is frame-based. Control structures and search strategies of these knowledge-based approaches will be discussed in the next section, together with the other publications.

3.2 CONTROL STRATEGIES FOR ROAD EXTRACTION

Because the control strategy influences the interpretation process, reviewed publications are categorized by the different control strategy as presented in section 2.3. The first categories are ways of hierarchical control: traditional bottom-up control, top-down control, in practice guided by a map or a human operator, and hybrid control, which integrates the previous approaches. The last category is heterarchical control, in particular the blackboard approach. For each of these categories low and high level image processing techniques will be discussed which are

used in the reviewed publications to build a road interpretation process.
A general characteristic of all road extraction procedures described in the reviewed literature is that they run in a monoplotting mode, i.e. use a single aerial image.

Authors	Landscape type			
	urban	suburban	rural	uncultivated
Airault et al. [1994]		X	X	
Bajcsy and Tavakoli [1976]			X	
Cleynenbreugel et al. [1990]				X
Fischler et al. [1981]				X
Fua and Leclerc [1990]		X		
Garnesson et al. [1990]		X	X	
Groch [1982]			X	
Grün and Li [1994]	X	X	X	
Gunst et al. [1991]			X	
Heipke et al. [1994]			X	
Hwang et al. [1986]		X		
Lemmens et al. [1988]	X			
Maillard and Cavayas [1989]			X	X
McKeown and Denlinger [1988]	X	X	X	
Nagao and Matsuyama [1980]	X	X	X	
Sijmons [1987]			X	
Stilla and Hajdu [1994]	X			
Ton et al. [1989]			X	
Vosselman and Knecht [1994]			X	
Wang et al. [1992]	X			
Zhu and Yeh [1986]		X	X	

Table 3.3 *Types of landscape in test images of reviewed publications*

3.2.1 BOTTOM-UP CONTROL IN ROAD EXTRACTION

Especially in early publications on road extraction a bottom-up approach is often used, since it overlaps with the traditional three step paradigm for image analysis (section 2.1.1). The aim is to achieve fully automatic road extraction. Table 3.4 summarizes the work discussed in this section arranged in chronological order.

Except for [Lemmens et al., 1988] most of them have satellite images or low resolution aerial images as input. Extraction of roads at this resolution is primarily approached as a problem of delineation of linear features. Characteristic for bottom-up road extraction is a sequential procedure of line detection followed by thresholding and finally noise removal and linking pixels in the binary image [e.g Bajcsy/Tavakoli 1976, Ton et al. 1989, Wang et al. 1992]. Since properties of individual pixels are used instead of more global properties, the result looks more like a collection of road pixels than like a logical road network. Dynamic programming offers the possibility to obtain a more global solution based on grey value pixels within a region of interest [Fischler et al. 1981, Lemmens et al. 1988].
A feature of bottom-up road extraction is that the whole image is in the same state of processing. A drawback of such a procedure is that the quality of processing steps heavily depends on each other, since it is a one-direction single track process. Fischler et al. [1981] partly handle this problem by combining the output of several edge detectors in a cost array and Sijmons [1987] combines the output of two parallel processes with a logical "or".
In general only few knowledge about roads is used, mainly based on radiometric properties and the linear shape of roads. None of it is implicitly included. It should be noted that values of radiometric properties needed for thresholding differ more or less for every image.

Author	Summary
Bajcsy and Tavakoli [1976]	structural thresholding on Landsat MSS followed by linking and noise removal in the binary image
Fischler et al. [1981]	minimum cost path determination in a cost array produced by various edge detectors on low resolution aerial images
Sijmons [1987]	median filtered image subtracted from original Landsat TM image combined with thresholding based on spectral properties
Lemmens et al. [1988]	comparison between two pairs of edge and line detectors on high resolution aerial images and road recognition by shape analysis of regions.
Ton et al. [1989]	developed their own operator to detect lines which are subsequently classified with the aim to detect minor roads in Landsat TM images
Wang et al. [1992]	analysis of a list of pixels along the gradient direction with additional thinning algorithms on SPOT images

***Table 3.4** Reviewed literature on road extraction using a bottom-up approach*

3.2.2 TOP-DOWN ROAD EXTRACTION BY UTILIZATION OF MAPS

Although the use of geographic information as additional knowledge source for aerial image interpretation has been widely recognized [Matsuyama, 1987], only a few publications make use of maps. Maps could be used to improve and classify the road network extracted from an image, however most publications [e.g. Roux et al. 1990, Li et al. 1992] do not go beyond the purpose of matching the road networks from map and image.

In this section publications are reviewed in which maps are used to constrain where to look and what to look for during interpretation of an image. In this case the map is used as a means to control road extraction in a top-down fashion. This is called map-guidance. One possible approach for map guidance is to search roads in the image with radiometric properties similar to those of roads present in the overlayed map. However, often newly constructed roads do not have equivalent radiometric properties, so no successful results are reported. [Maillard/Cavayas, 1989], [Cleynenbreugel et al., 1990] and [Stilla/Hajdu, 1994] are examples of publications in which other map-guided approaches are used. Maillard and Cavayas [1989] use the road map as a logical framework to search new roads, which are usually connected to existing roads. Roads existing in the map are tracked on the image to search intersections with new roads. Next new roads are tracked based on previously identified intersections. In [Cleynenbreugel, 1991], which contains more details than [Cleynenbreugel et al., 1990], it became evident that existing roads are not verified, but used as a logical framework when establishing relationships between extracted line pieces. Apart from a road map, also a DTM is used to verify constraints on maximum allowed slope for connectable line pieces. Stilla and Hajdu [1994] describe the contents of the map in a so-called image description graph, which is input for knowledge-based image analysis. Based on the knowledge described in this graph expectations are defined for attribute values of objects in the image, like position and width of roads. The knowledge from the map does not influence the result of image analysis, but rather the processing sequence in which objects are searched dependent of their presence in the map. They state that map knowledge reduces the processing time significantly. If the map is outdated and needs to be updated, processing time will probably not reduce since expectations for changes are not defined.

Authors	Summary
Maillard and Cavayas [1989]	map is used as logical framework to extract roads from SPOT images
Cleynenbreugel et al. [1990]	expert system which uses GIS information to verify roads extracted from SPOT images
Stilla and Hajdu [1994]	use maps to create a knowledge base from which expectations about roads in aerial images of urban scenes are inferred

***Table 3.5** Reviewed literature on road extraction using a map-guided approach*

3.2.3 TOP-DOWN ROAD EXTRACTION BY HUMAN INTERACTION

The purpose of semi-automatic road extraction approaches is to develop techniques, which can be used to speed up manual tasks in map production, since full automation is not realistic in the near future. In all reviewed papers about semi-automatic road extraction, summarized in table 3.6 in alphabetical order, interaction by a human operator is used to initiate the search for roads in a top-down fashion. Although human interaction could also be used to evaluate extracted roads or to classify the road, no examples of this approach were found.

Authors	Summary
Airault et al. [1994]	extrapolation from one manually indicated point in the direction with most homogeneous surface
Fua and Leclerc [1990]	deformation of a manual indicated polygon using snakes
Gunst et al. [1991]	linear extrapolation from road segment by combination of profile analysis and dynamic programming
Grün and Li [1994]	interpolation between a set of manual indicated points by dynamic programming
Heipke et al. [1994]	road segment extension by edge-based road tracking in a search window
Vosselman and Knecht [1995]	extrapolation from road segment using a Kalman filter and evaluation of the grey level profile

Table 3.6 *Reviewed literature on road extraction using a semi-automatic approach*

When using human interaction to initiate road extraction, the control strategy can be characterized as top-down, since starting from the hypothesis that the indicated segments can be classified as roads, the exact locations are determined using the low level image data. As mentioned in appendix A, basically there are two approaches in interactive initiated road extraction:
1. Extrapolation
2. Interpolation

Interactive road extraction strategies based upon extrapolation extend a small road segment, which a human operator indicates manually. Road trackers are applied to automatically extend the small segment. Variations upon common road trackers based on profile analysis (see in section 3.2.4 e.g. [Groch, 1982], [McKeown/Denlinger, 1988]) are also suitable to apply semi-automatically. In [Gunst et al., 1991] dynamic programming is used to adjust the road point predicted by such a profile analyser if no characteristic profile was found. Vosselman and Knecht [1995] use Kalman filtering to provide the profile analyser with predictions of the position of the road. In [Airault et al., 1994] only one point on the road is indicated manually from which paths with varying length and direction are hypothesized and next evaluated using a criterion for homogeneity of the road surface. In [Heipke et al., 1994] the human operator provides the road tracker with a starting point and an initial direction. This road tracker is based on edge detection and edge following within a small window, related to the search direction. This tracker yields the road side whereas the other trackers are used to determine the axis of the road. However, trackers based on profile analysis could produce the road side as well, if the human operator indicates it.

Other strategies interpolate between a set of sparse points distributed along the complete road which the human operator indicated to roughly describe the road. Dynamic programming is used by Grün et al. [1994] to determine semi-automatically the exact location of the road by maximizing a merit function. Fua and Leclerc [1990] use active contour models or snakes which are deformed with the aim to minimize the total energy, which is a function of deformation and photometric energy.

3.2.4 HYBRID CONTROL IN ROAD EXTRACTION

Reviewed literature on road extraction using hybrid control, which integrates bottom-up and top-down processing, is summarized in table 3.7. The purpose of Hwang et al. [1986] is to develop a general control strategy, which is tested on the case of interpretation of aerial images, while the other authors design a procedure suitable for the detection and tracking of linear features in general [Groch, 1982] and more [McKeown/Denlinger, 1988] or less [Zhu/Yeh, 1986] for roads in particular. Knowledge is represented explicitly by [Hwang et al., 1986] and [Zhu/Yeh, 1986], but incorporated in the procedure of Groch [1982] and McKeown and Denlinger [1988].

Authors	Summary
Groch [1982]	starting point detection and road tracking by analysis of perpendicular profiles
McKeown and Denlinger [1988]	road tracking by profile analysis and edge linking together with detection of road features
Zhu and Yeh [1986]	criteria for edge-based seed selection, growing and bridging gaps are expressed by production rules
Hwang et al. [1986]	bottom-up segmentation and top-down reasoning about missing road parts using geometric and spatial properties represented in frames

***Table 3.7** Reviewed literature on road extraction using hybrid control*

[Groch, 1982] and [McKeown/Denlinger, 1988] both describe systems based on road trackers, which sequentially extend roads by prediction and verification of parts of them. This method of working justifies to define it as top-down control, even though the models of the road used for prediction are not explicitly represented. The road tracker of Groch [1982] verifies predicted road points by analysis of the characteristic grey level profile perpendicular to the predicted direction of each road. In [McKeown/Denlinger, 1988] a sophisticated tracker based on profile analysis as well as an edge-based tracker can be activated by high level routines. Detection of road features, like road width changes, intersections, or vehicles, causes adaptation of parameters for prediction and verification. Gaps are bridged while extending road pieces, independently of other road pieces.

Road trackers are suitable for application in semi-automatic road extraction systems (see section 3.2.4), but when combined with detection of starting points, they operate fully automatic. Groch [1982] describes a starting points detector based on analysis of grey-value profile along two circular sample lines around each candidate point at the crossing of grid lines. McKeown and Denlinger [1988] refer to a starting point detector described in [Aviad/Carnine, 1988] and later also in [Zlotnick/Carnine, 1993]. This detector forms trajectories of pixels which lie in the middle between two anti-parallel edges. Since the described detectors of starting points proceed bottom-up, the complete road extraction system can be considered as driven by hybrid control.

[Hwang et al., 1986] and [Zhu/Yeh, 1986] both describe a knowledge-based system which integrates bottom-up and top-down control. [Hwang et al., 1986] describes SIGMA, already discussed in section 2.5.3 as an example system which uses frames to represent knowledge. When no specific goal is given, processing starts with activation of a simple bottom-up segmentation based on grey-value thresholding which produces regions. To select candidate road pieces and houses, criteria for shape and contrast expressed in frames are evaluated. Road pieces are extended and linked, eventually by re-segmentation with another method within a small window at a predicted location. Also the spatial relationships between houses and roads are used, for example to search driveways. A drawback is that the segmentation methods are too simple to yield results accurate enough for map production. Zhu and Yeh [1986] use the anti-parallelism of edges to segment road pieces, with additional geometric and radiometric properties to select reliable road pieces. Gaps are bridged and road pieces extended by evaluation of edges within a small window. The consequences of various combinations of four criteria for the quality of edges are expressed explicitly in production rules.

3.2.5 Heterarchical control in road extraction

In both [Nagao/Matsuyama, 1980] and [Garnesson et al., 1990] the systems for knowledge-based road extraction are based on a blackboard architecture, which is a way of heterarchical control. However, the blackboard of Nagao and Matsuyama [1980] aims to recognize all regions, extracted based on their spectral properties, while the blackboard of Garnesson et al. [1990] aims to activate only these procedures which extract objects with a high probability to be present. There discriminating features are summarized in table 3.8 and fig. 3.1 and 3.2 depict both blackboard configurations.

Nagao and Matsuyama [1980] start with an initial smoothing and segmentation. The produced regions are stored on the blackboard. The blackboard can activate two types of knowledge sources:

1. characteristic region extractors;
2. object-detection subsystems.

Several characteristic regions are first identified, like elongated, homogeneous and shadow regions. The subsystem for road detection focuses its attention on elongated regions and applies specialized analysis methods such as connection of candidate regions only in those focused areas. The control system tries to correct segmentation errors of irregularly shape regions by split and merge and solves conflicts if a certain region is recognized as multiple different objects.

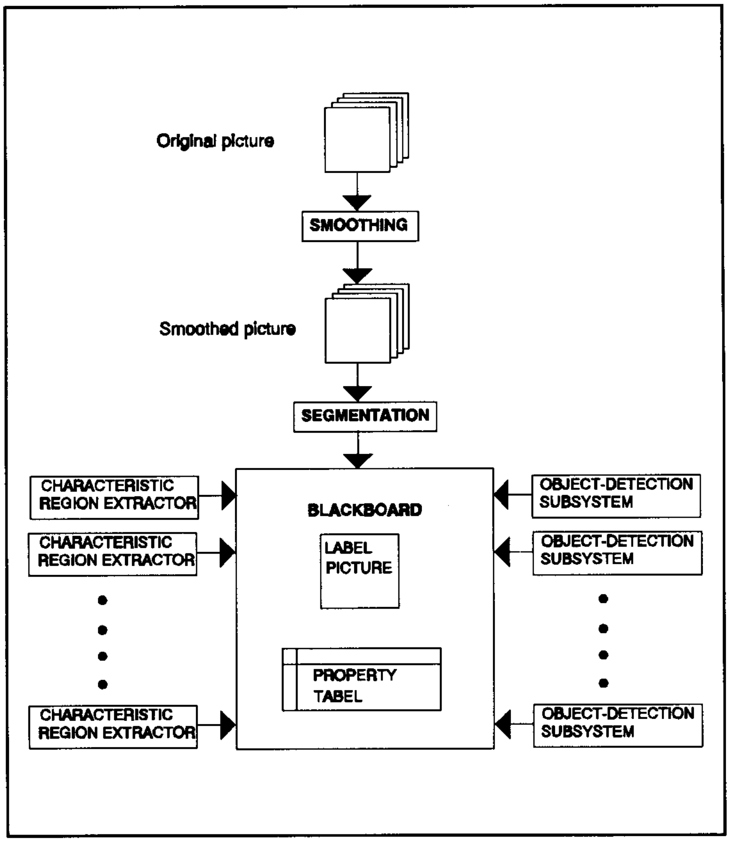

Fig. 3.1 Blackboard configuration from [Nagao/Matsuyama, 1980]

Garnesson et al. [1990] present MESSIE: Multi Expert System for Scene Interpretation and Evaluation. More details about MESSIE can be found in [Garnesson, 1991].

The blackboard can activate two types of knowledge bases:
1. Specialists, operating at the low level of processing to extract objects from the image;
2. Controllers, operating at the high level of processing in order to control the interpretation strategy.

The strategy is to start searching the most salient objects in the scene, because it is assumed that they can be recognized most reliably. If one or more objects are found, the controller of location uses knowledge about the context to generate hypotheses about the presence of a related object within a defined search area. The supervisor chooses which specialists should be activated to segment the hypothesized object. There is a separate specialist for the detection of roads, which constructs boxes from anti-parallel edges. Hypotheses are evaluated by the scene controller. Only conflicts which lead to identification of the same region as two different objects are handled by MESSIE.

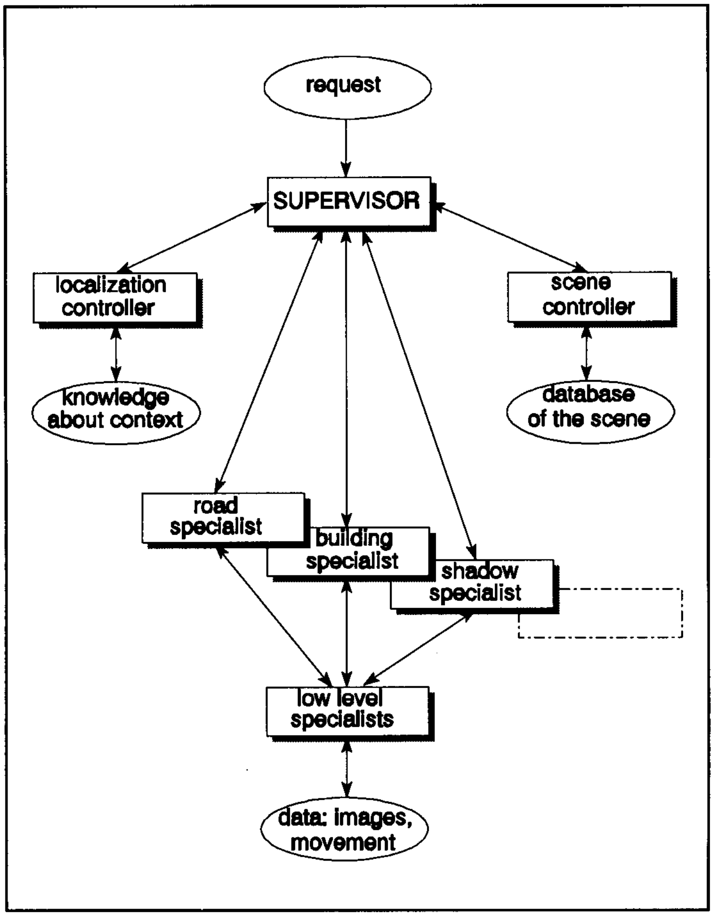

Fig. 3.2 Blackboard configuration from [Garnesson, 1991]

Author	Summary
Nagao and Matsuyama [1980]	characteristic regions are input for object-detection subsystem
Garnesson et al. [1990]	after salient objects are found, controller generates hypotheses about other objects in their context and activates specialists

Table 3.8 Reviewed literature on road extraction using a blackboard approach

3.3 ROAD CHARACTERISTICS

Because in this thesis the contents of the knowledge base will be emphasized, an overview is given of the characteristics of roads used for road extraction in previous work. The next subsections will list which properties are used to describe roads and, if they are used implicitly, how these properties were used for road extraction in various publications. One should notice the different functions of the properties: to search roads (inference), or to recognize extracted primitives as roads (validation). In order to structure the description of road properties used in the reviewed literature, we use the division defined by Garnesson et al. [1990] into four concepts:

1. geometry, includes size and shape of road elements;
2. radiometry, includes grey value and colour of road elements and features determined from the grey values in the image, like texture;
3. context, interaction of roads with other objects in the scene, including spatial relationships;
4. function, direct consequences of the function of road networks.

The last two concepts are most difficult to apply for road extraction. When a map is used for road extraction the a priori road description is often limited or even absent, because extracted properties of lines or regions are compared with properties in the map.

3.3.1 ROAD-SPECIFIC GEOMETRIC PROPERTIES

Geometric properties are related to the coordinate description of the road. Therefore they differ for lines and regions. Line-like properties can also be used for axes of regions.

a) Roads have constant width
Zhu and Yeh [1986] and Garnesson et al. [1990] both search candidate road pieces by trying to find two parallel edges, thus separated by a constant distance, and use this width for road growing. The grey level profile depends on the width of the road and for that reason this property is implicitly included in road trackers based on profile analysis. When bridging gaps between to road regions Nagao and Matsuyama [1980] check if their widths are about the same. Fua and Leclerc [1990] include an additional term in the deformation energy, which enforces the parallel constraint, defined as a minimization of the squared sum of width differences.

b) The width of the road is bounded at a certain resolution
Bajcsy and Tavakoli [1976] use the assumption that a road is one, two or three pixels wide in a Landsat image to select road pieces. Lemmens et al. [1988] select regions with a width between 2 and 12 metre on the ground.

c) Isolated roads have a minimal length
This property is used to justify removal of short segments in [Bajcsy/Tavakoli, 1976] and [Wang et al., 1989].

d) The path of the road is usually smooth and does not have small wiggles.
This property applies on resolutions where the road corresponds to a line, but also for the axis of the road on higher resolutions. Groch [1982] and Gunst et al. [1991] model the path locally by a polygon and McKeown and Denlinger [1988] by a parabola. Airault et al. [1994] predict several paths from one point by polygons with multiple sides of variable length and direction. Vosselman and Knecht [1994] use a Kalman filter for prediction of the next point of the road during tracking to smoothen the path. Grün and Li [1994] include in their algorithm for dynamic programming the property that the second derivatives of path coordinates attain a minimum to prevent small wiggles. For that reason Lemmens et al. [1988] include minus the cumulated differences in the direction along the path in their maximization cost function for dynamic programming. Ton et al. [1989] use the same measure for curvature together with the length of road segments between junctions to discriminate major roads from minor roads.

e) The side of the road is fairly straight
The straightness of edges is often included in processing applied in combination with edge detection. Approximation of a chain of pixels by polygon is done by Garnesson et al. [1990], Heipke et al. [1994] and Wang et al. [1989] resulting in a raster to vector conversion. McKeown and Denlinger [1988] model edges, just like the road axis, locally by a parabola. Zhu and Yeh [1986] indicate straightness by the second moment of the chain code, used to describe the edge and classify it into high, medium and low. Fua and Leclerc [1990] approximate the edge by a polygonal curve. Deformation energy is defined by the curvature in the energy minimization function.

f) Roads are elongated regions
Lemmens et al. [1988] use elongatedness, defined as length divided by width, as one of the criterion for road regions. Nagao and Matsuyama [1980] and Hwang et al. [1986] use elongatedness as one of the properties to select candidate road regions and calculate it by dividing the length and width of the region's bounding rectangle.

3.3.2 ROAD-SPECIFIC RADIOMETRIC PROPERTIES

Properties directly based on grey values in black-and-white or colour images, as well as properties of grey value differences used to extract edges, are both considered to be based on radiometry.

a) Roads are build from materials, for example concrete and asphalt, that have specific spectral properties
Bajcsy and Tavakoli [1976] determine a range of grey values to threshold to image, based on spectral properties of concrete and asphalt. The spectral property of asphalt and concrete is that they correspond to high grey values in the red, green and blue band. Sijmons [1987] selects pixels with among others high grey values in these spectral bands of a Landsat TM image.

b) Road materials do not vary much and their spectral properties are similar within a short distance
Grün and Li [1994] translate this property in their road model for dynamic programming by the constraint that grey value differences along the road attain a minimum. Airault et al. [1994] choose a path which optimizes a surface homogeneity criterion: the minimal grey level variance of pixels along the path. Nagao and Matsuyama [1980], Zhu and Yeh [1986] and Garnesson et al. [1990] use this property during road extension by comparing the mean grey value of a part of the road with the successive part before connecting them. Lemmens et al. [1988] merge adjacent regions if the difference between their mean grey values in the three spectral bands is below a certain threshold, set to 20.

c) Roads for the most part show significant contrast with their surroundings
Bajcsy and Tavakoli [1976] use a predefined value for the grey level difference to select road pieces. Hwang et al. [1986] use a measure for contrast explicitly as one of the properties to select candidate road regions. This property also leads to the fact that roads usually have significant edges. Zhu and Yeh [1986] use the average magnitude of the edges and the number of edge pixels below a threshold value for the magnitude as measure for the quality of the side of a road. Fua and Leclerc [1990] use minus the average edge magnitude to represent photometric energy, component of the total energy which is minimized. Airault et al. [1994] detect the edges to adjust the axes during road tracking by surface homogeneity.

d) Roads are light-coloured linear features on a dark background
This property leads to the fact that edges at opposite sides of a consistent region have opposite directions and can be taken together as anti-parallel pairs. Zhu and Yeh [1986], Aviad and Carnine [1988], Zlotnick and Carnine [1993] and Garnesson et al. [1990] all detect anti-parallel edges to extract roads. This property is also implicitly assumed when applying a line detector like in [Ton et al., 1989] or [Wang et al., 1992]. For determination of the maximum cost path by Fischler [1981], Lemmens [1988], Grün and Li [1994] and Gunst et al. [1991] the road model includes that the sum of the grey values (or their second derivatives in the direction perpendicular to the road) along the path attain a maximum, which is based on this property.

e) Roads have a characteristic grey level profile
This feature is the basis for all road trackers based on profile analysis. Since this profile is characteristic, but different for every road, an initial profile model is obtained from manually indicated or from automatically detected road pieces. Several measures are used to evaluate if an extracted profile has this characteristic shape. Groch [1982] codes the shape of the profile and uses this code for comparison. McKeown and Denlinger [1988] use the cross-correlation as criterion. Gunst et al. [1991] use weighted root mean square difference and Vosselman and Knecht [1994] use least squares matching in addition to the cross-correlation.

f) The grey level profile changes either very gradually or suddenly
To cope with gradual changes, McKeown and Denlinger [1988] and Gunst et al. [1991] update the profile model by calculating a weighted average between the old model and the extracted profile, Vosselman and Knecht [1994] do not, since this leaded to a bias in the estimated road

position. Only McKeown and Denlinger [1988] detect sudden changes by anomaly detection and activate in this case the edge-based tracker.

3.3.3 Contextual information

If relationships of roads with other types of objects, in particular cars, houses, and vegetation, are included, this is considered as contextual information.

a) Cars drive on the road
Garnesson [1991] uses the previously detected road as search area to extract cars with the car specialist. Identification of cars increases the confidence that the detected area is really a road.

b) Absence of vegetation on the road
Sijmons [1987] uses this property as one of the criteria when selecting pixels with a low Green Vegetation Index, a measure for the amount of vegetation. Nagao and Matsuyama [1980] bridge gaps between road regions only if the region in between does not have the spectral properties of vegetation.

c) Roads are usually parallel to a row of houses
Hwang et al. [1986] show how previously detected houses can be used to define a search area where to expect a road.

d) Houses and roads are usually connected by a driveway
In suburban scenes Hwang et al. [1986] use this property to define a search area where to expect a driveway.

3.3.4 Functional features

The most important function of a road is its function as a means of communication between different locations such as cities, buildings, parks, rivers and lakes. This leads to the next properties:

a) Roads are connected to form a network.
Widely used methods to provide connection and construct junctions are linking of line pieces, assumed to be part of the road side or axis, if their distance and difference in direction is below a threshold [e.g. Bajcsy/Tavakoli 1976, Aviad/Carnine 1988, Ton et al., 1989] and bridging gaps between regions, assumed to be part of the road surface, using geometric as well as radiometric properties [e.g. Nagao/Matsuyama 1980, Hwang et al., 1986, Zhu/Yeh 1986]. Lemmens et al. [1988] choose 25 metre on the ground as a maximum to bridge gaps. Ton et al. [1989] require that the length of the segments on both side of the gap is at least three times the length of the gap. Cleynenbreugel et al. [1990] define that for particular types of images the roads intersect perpendicular to form a grid pattern and use this fact to construct L-shaped, U-shaped and

rectangular structures from road segments.

b) The slope along the track of the road has a maximum to allow cars to drive uphill.
Cleynenbreugel et al. [1990] use a constraint on the maximum tolerable slope, calculated from an accompanied DTM, to verify line elements in mountainous terrain.

c) The local change in direction of the road is upward bounded from traffic flow requirements.
Grün and Li [1994] include in their road model for dynamic programming an upper bound for the local curvature of the road. In [Heipke et al., 1994] edge tracking stops if the difference in direction between successive edges is larger than a certain threshold value. Lemmens et al. [1988] choose 0.3 rad. as threshold for the maximum difference in direction between the principle axes of regions. The change in direction of the road during profile analysis is limited by the width difference between the profile model and extracted profile. In [Gunst et al., 1991] these widths are chosen such that the change in direction of the road is maximal 15 degrees. Aviad and Carnine [1988] split axis after detection, if a sudden change in direction appears. In [Vosselman/Knecht, 1994] the maximum road curvature is used to derive the system noise of the Kalman filter.

3.4 RESULTS AND DISCUSSION

Roads in uncultivated areas are only extracted from low resolution imagery (see tables 3.1, 3.2 and 3.3). This is also the most interesting application, since especially for those areas in for example developing countries few up-to-date maps will be available. From rural and (sub)urban areas usually maps with more detail are made and consequently large-scale imagery is required. However, in most results roads are represented by a one-pixel thick line, in larger scale images corresponding to the axis of the road. Exceptions are [Nagao/Matsuyama, 1980] in which roads are represented by elongated regions and Hwang et al. [1986], Zhu and Yeh [1986], Fua and Leclerc [1990] and Stilla and Hajdu [1994] who detect parallel edges. Viewing road elements as part of a stable network already during segmentation is only done by Maillard and Cavayas [1989], while only McKeown and Denlinger [1988] recognize road features like intersections, overpasses, surface material changes and road width changes.
As already noted by Cleynenbreugel [1991] validation of results is difficult, as well during road extraction as afterwards. Because of the absence of an exact road model, roads of which the appearance is not included in the used road model, will not be found. Segmentation is always the weakest step in the interpretation process and errors in segmentation propagate into recognition. Results show that a combination of several road extractors performs better than either extractor alone [e.g. McKeown/Denlinger 1988], because of the complex contents of aerial photographs. Evaluation of the result by comparing it with a map or manual road extraction is only done for bottom-up approaches. A problem is how to compare them. Statistical measures like the percentage correctly classified pixels do not represent the practical suitability of a method, because they do not make clear how much editing work is needed afterwards to produce a map. Accuracy analysis is only done by Lemmens et al. [1988].

The number of test images used in a publication is a nice indicator for practical suitability of its developed road extraction strategy. Fig. 3.3 shows in a histogram how many reviewed publications use a certain number of test images. It shows clearly that in most cases only one or two test images were used. The low number of test images makes it likely that many results are obtained by trial and error. However, in most publications no report is made about a test procedure to tune parameters. [McKeown/Denlinger, 1988] is the only publication which describes an extensive test procedure on many images. They used 35 roads in 26 training images to tune parameters and another 35 roads in 18 images for testing. Besides they compare profile analysis and edge tracking as single and combined methods. Because of this extensive test procedure, this method is considered to be promising for our application.

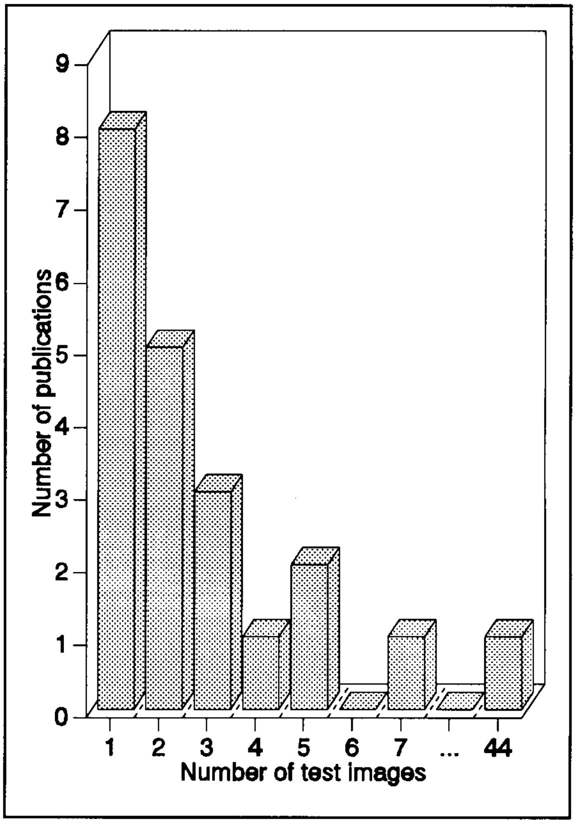

Fig. 3.3 Number of publications using a certain number test images

Semi-automatic techniques are mainly applied on suburban or rural scenes. Because results of interactive techniques coincide best with results a human operator achieves, these techniques are most suitable for practical application. One reason is that they concentrate on relatively the easiest part: local extraction of road elements between junctions. The more difficult part, reasoning about validation of extraction and construction of a road network, is done by a human operator. Especially these tasks will require much attention during further automation. Results show that knowledge-based techniques are most suitable. They are mainly applied on large-scale imagery.
It is striking that sophisticated reasoning like in [Hwang et al., 1986] is often combined with simple image processing techniques like grey level thresholding whereas in case of more advanced image processing techniques, like dynamic programming by Fischler et al. [1981], weak mechanisms are provided to validate and rectify results by reasoning. The requirements that combination of segmentation techniques and correction of segmentation errors put on the control and reasoning strategy will be discussed in the chapter 4.

A lot of used road properties are very general and not very road-specific, but more line-specific. Spatial relationships are only included by Hwang et al. [1986] and Garnesson et al. [1990]. Most advanced properties are implicitly included, also in knowledge-based approaches. Simple properties like elongatedness and width are most often represented explicitly. Because functional properties are poorly present, also distinction between functional parts of the road network, like motorways versus dirt tracks and fly-overs versus crossings, is never made. Requirements for the contents of the knowledge base will be defined in the next chapter.

Only few applications use maps as knowledge source and even fewer use maps to guide the interpretation. Cleynenbreugel et al. [1990] use maps to validate detected roads in contrast to Maillard and Cavayas [1989] and Stilla and Hajdu [1994] who use information in the maps to search new roads. Possible changes between the outdated map and the aerial image are not considered and consequently not explicitly modelled. This will be one of the contributions of the map-guided interpretation strategy designed in the next chapter.

CHAPTER 4

CONCEPTS FOR KNOWLEDGE-BASED ROAD EXTRACTION

In chapter 1 it is argued why knowledge-based techniques should be used for interpretation of aerial images. The goal of this thesis is defined as design of a strategy for map-guided interpretation of road networks in aerial images. In this chapter the concepts for the designed strategy will be discussed, which are also briefly described in [Gunst/Hartog, 1994]. Based on the reviewed literature in chapter 2 and 3 requirements for the interpretation strategy, contents of the knowledge base and utilization of maps will be defined. Next, concepts for the designed road model and interpretation strategy will be outlined. In order to illustrate the concepts, finally examples of their application on large scale photographs will be given.

4.1 REQUIREMENTS FOR INTERPRETATION OF AERIAL IMAGES

The concept of knowledge-based interpretation is to construct an interpretation strategy of low level image processing and high level reasoning to search objects. Comparison of symbolic descriptions of objects extracted from the image with a priori object knowledge yields recognition and may influence the interpretation strategy. This definition shows that choices, which need to be made based on the reviewed literature in chapter 2 and 3, mainly concern the next subjects:

1. Interpretation strategy:
   - control of the process (4.1.1);
   - low level processing (4.1.2);
   - high level reasoning (4.1.3).
2. Contents of knowledge base:
   - types of knowledge (4.1.4);
   - representation formalism (4.1.5).

4.1.1 CONTROL STRATEGY

Because aerial images contain many objects with large variations within each object class, the sequence of processing steps cannot be determined reliably beforehand. Consequently the traditional bottom-up approach of low level segmentation of the whole image followed by high level reasoning [e.g. Bajcsy/Tavakoli 1976, Wang et al. 1992] yields many omission and commission errors. The main reason is that especially segmentation is often weak and processing steps heavily depend on each other. Blackboard systems and integration of bottom-up and top-down control both offer possibilities for reprocessing to deal with this problem and yield more promising results [e.g. Hwang et al. 1986, Garnesson et al. 1990]. A merit of the blackboard model [Matsuyama, 1987] is that due to its modular system organisation, object detection modules can easily be appended. However, knowledge about objects is embedded in the programs and is as a consequence more difficult to modify. In blackboard systems processes are activated when certain conditions of the data are fulfilled. As blackboard systems become more complex, it becomes harder to keep the interpretation process under control. In this case sophisticated control structures are required [Nagao, 1984]. Because this is intrinsic to the concept of heterarchical control which underlies the blackboard model, hierarchical control in the sense of integrating bottom-up and top-down processing is favoured.

The importance of combining both bottom-up processing with top-down reasoning for complex images like aerial photographs has been widely recognized [Matsuyama, 1987]. In contrast to [Cleynenbreugel et al., 1990] top-down control initiated by objects from the digital map is proposed for initial processing. The advantage of initial top-down control is that only parts of the road network are extracted from the image which are consistent with the road network in the digital map. Accordingly the main drawback of initial bottom-up processing is that it is very difficult to obtain a consistent road network from segmented road elements. Some of the problems are that the result may depend on the sequence in which initial road elements are processed and that many non-road elements are included in the reasoning process. Often [e.g. Cleynenbreugel et al. 1990, Stilla/Hajdu 1994] the control works on the symbolic level where it reasons with extracted line elements and regions, without including the image data itself. This drawback is handled in [Nicolin/Gabler, 1987] by including a feedback loop to extract missing objects from the image, based on general properties of regions, before classification into real-world objects is performed.

In conclusion hybrid control with initial top-down processing is proposed for road extraction.

4.1.2 Low level image processing

Image processing techniques can be used as tools to analyse the image. However, it has been shown that none of the image processing operators are perfect and that we have to select useful image processing operators and combine them into effective image analysis processes [Matsuyama, 1987]. One of the problems is how to integrate them. When the output of several low level processing techniques is combined logically [e.g. Sijmons, 1987] or numerical [e.g. Fischler et al., 1981], no optimal use is made of strength and weakness of a particular method for a certain situation. This is also the case if during road following two trackers are alternated, like in [Gunst et al., 1991] and [McKeown/Denlinger, 1988]. Therefore goal-directed segmentation was introduced. This means that a segmentation technique is performed depending on the goal in a top-down object extraction process. By generation of multiple goals, combination of several image processing techniques is realized. Nazif and Levine [1984] and Hwang et al. [1986] propose an expert system which uses knowledge about image processing techniques and the expected appearance of an object to select the most promising low level processing technique from a pool. However, it is very difficult to define characteristics of image processing techniques, which can be used as criteria for selection. Research on this topic is still going on [e.g. Clément/Thonnat, 1992], but is not within the scope of this thesis. The approach of

Garnesson et al. [1990] who define specialized segmentators for specific objects, like houses and roads, is more suitable for this work. However, the characteristics of a specific segmentator not only depend on the type of object searched, but also on the relationship with the object from which it is searched. For example: texture characteristics of a road can be included in a segmentation algorithm to extract another road nearby, but the fact that roads often run along a row of houses can be used to define a region of interest in which another segmentation algorithm is applied. Therefore in the designed interpretation strategy there is a specific segmentator for each relationship. A requirement for design is that it should be easy to replace one image processing technique by another, for example a semi-automatic technique.

Beside selection and combination of low level image processing techniques there is also the problem of determining appropriate values for parameters. Wrong parameters can cause errors in interpretation. Another problem is related to the fact that goal-directed segmentation offers the possibility to reduce the huge amount of computation at image level by defining a restricted part of the image in which the technique is applied. The shape and location of this region of interest can also be subject to errors. In order to handle these problems it is required that it is possible to define alternatives for the region of interest, the image processing technique, and its parameters.

4.1.3 High level reasoning

The aim of high level reasoning in a hybrid control strategy is to recognize objects in the result of low level processing and to generate goals for segmentation of other objects. Therefore it is natural to use an object-oriented fashion of reasoning. An object is an entity which takes the central position in the reasoning process. From a software engineering perspective object-oriented reasoning provides the best solution for a modularity and thus allows extension of the interpretation strategy with other topographic objects than road networks as was required in section 1.3. From a model building point of view object-oriented reasoning offers the opportunity to represent directly physical objects and their characteristics in the knowledge-base, which provides more insight into the process of image understanding. Object-oriented reasoning is used for road extraction in [Cleynenbreugel et al., 1990], [Garnesson et al., 1990], [Hwang et al., 1986] in contrast to for example [Zhu/Yeh, 1986] and [McKeown/Denlinger, 1988] who generate processing tasks at the high level.

Since errors are inevitable in image analysis, verification of low level image processing results as well as high level control are required. High level reasoning needs to decide when an alternative segmentation technique, alternative parameters or an adapted region of interest are used. Besides it needs to detect and solve inconsistencies in the classification of identified objects, e.g. based on their context. An example of such an inconsistency is the detection of a level road junction on a motorway. Verification of results is difficult because of the absence of an exact road model.

4.1.4 TYPES OF KNOWLEDGE

The types of knowledge included in the road model should reflect the complex contents of aerial images as specified in section 1.2.1.

Because the density of objects in aerial images is very high, semantic information is necessary to identify meaningful individual objects, rather than syntactic information to extract structures. However, in previous work road extraction is primarily approached as a problem of linear feature detection and grouping of lines to form a network. Consequently the work on road extraction is mainly based on three knowledge sources:

- geometrical properties, concerning width and curvature of roads;
- radiometric statistics, based on contrast and homogeneity of intensity, represented by grey values in the image;
- grouping rules, based on constraints like collinearity, parallelism and perpendicularity.

As a result roads cannot be recognized unambiguously, because these knowledge sources are too simple to distinguish for example roads from railways. Therefore it is required that employed knowledge sources are based on a more strict model, which explains the possibility to find particular parts of the road network and uses intrinsic properties of roads for identification. For that reason standards for road construction [Rijkswaterstaat, 1975] were studied and included in our road model. These standards often differ for specialized classes of roads, like motorways or streets, depending on their function. Therefore recognition of specialized classes is also profitable if only general classes need to be mapped, because objects can be identified more reliably with these specific object properties.

Beside knowledge concerning individual objects, the image contents also requires three types of relational knowledge. The complex composition of parts of the road networks requires inclusion of part-whole relationships. They can simplify recognition of complex objects. Specialization-generalization relationships are not only useful for definition of more specific object properties, but they also yield semantic constraints between objects. Motorways for example cross each other by fly-overs and not by level road junctions. The need of context to interpret images of complex object spaces, leads to the conclusion that spatial relationships need to be included. Spatial relationships can help to determine where to expect a certain object, but also yield constraints for the presence of objects.

Because of the object-oriented fashion of reasoning, procedural knowledge should be attached to the relationship between the known and searched object type (see section 4.1.2). It consists of knowledge about image processing techniques for goal-directed segmentation and requires that the relation between the searched object type and appropriate segmentation techniques is known.

4.1.5 KNOWLEDGE REPRESENTATION FORMALISM

The previous section shows that the knowledge which needs to be represented is enormously varied. Because declarative as well as procedural knowledge are involved, frames are the most suitable representation formalism. Frames also support object-oriented reasoning, as was required in section 4.1.3 and allow easy extension for other topographic objects. The underlying semantic network reflects the discrimination between object and relational knowledge.

4.2 REQUIREMENTS FOR UPDATING AND UTILIZATION OF MAPS

From the four subjects which Matsuyama [1987] notes as yielding problems when using maps for image understanding (section 1.3), within the scope of this thesis only requirements for the data structure to store maps and the design of a map-guided interpretation strategy will be discussed.

4.2.1 DATA STRUCTURE TO STORE MAPS

Digital map data needs to be stored during interpretation in a format which is accessible for image analysis. The easiest way is to transform it to a raster image like is done in [Maillard/-Cavayas, 1989]. However, useful vector data is lost, which could have been used to calculate accurate geometric attributes or to predict extended road elements. Cleynenbreugel et al. [1990] and Stilla and Hajdu [1994] symbolically describe the map by lines and meaningful compilations of lines like collinear and parallel structures. The same kind of primitives are extracted from the image and can be compared. However, such a description complicates returning to the image data for re-segmentation, using the previous result as approximate search area. Besides, some properties (e.g. area) can easier be calculated from raster data. In this case the availability of raster data would be advantageous. Therefore it is required that the digital map is stored in such a format that vector as well as raster data can easily be retrieved.

4.2.2 MAP GUIDANCE

The question is how outdated or incomplete maps can be utilized during interpretation of aerial images. Instead of using the map to verify segmentation results, like by establishing relations with existing roads in [Cleynenbreugel et al., 1990] or by comparing attributes in [Stilla/Hajdu, 1994], our aim is to utilize the map as well for segmentation, like in [Maillard/Cavayas, 1989], as for the detection of changes. The complexity of verifying whether a road segment from the map changes or not is shown in [Gunst/Lemmens, 1992]. Since the number of possible changes in outdated maps is limited, knowledge about possible changes in time should be used to detect changed and new objects in order to update an outdated map. Changes are required to be modelled explicitly in order to support maximum flexibility and modularity of the system. Also incomplete maps can be used to guide image interpretation. Incomplete maps can originate from another scale of mapping with less detail and more generalized object classes. The location of objects in incomplete maps creates a logical framework to search new objects while their properties can tailor a priori knowledge for this specific situation. In this way new objects will be found more reliable and more accurate. These objects can be used to update incomplete maps.

4.3 OBJECT-ORIENTED MODEL FOR ROAD NETWORKS

As was required in 4.1.4 the road model should reflect the complex contents of aerial images. The designed road model is restricted to motorways and other through roads and does not include small (unpaved) roads in towns and villages like streets.
Important for an object-oriented approach is the definition of relevant object classes. They should be chosen such that spatial, part-whole as well as specialization-generalization relations can be modelled. The representation of the road network differs considerably for different scales. At small scales, like satellite images, road networks can be represented by lines forming a network of intersections and roads. However at large scales the road surface should be represented in the digital map as a planar feature containing details like road markings. Therefore it was decided to discriminate between three levels of detail in the road models: small, medium and large scale. Levels are connected by part-whole relations. For example: an object at medium scale *consists of* several objects at large scale, while one of these objects at large scale is *part of* the object at medium scale. The complete part-whole hierarchy is shown in fig. 4.1. Similar road terms in fig. 4.1 are used in this thesis with a specific meaning, expressing at which scale a part of the road network is defined.
Appendix B explains and defines all road terms that are used in this thesis. Examples of the appearance of all objects in fig. 4.1 are given in fig. 4.2 as a zoom on the same situation at various scales.

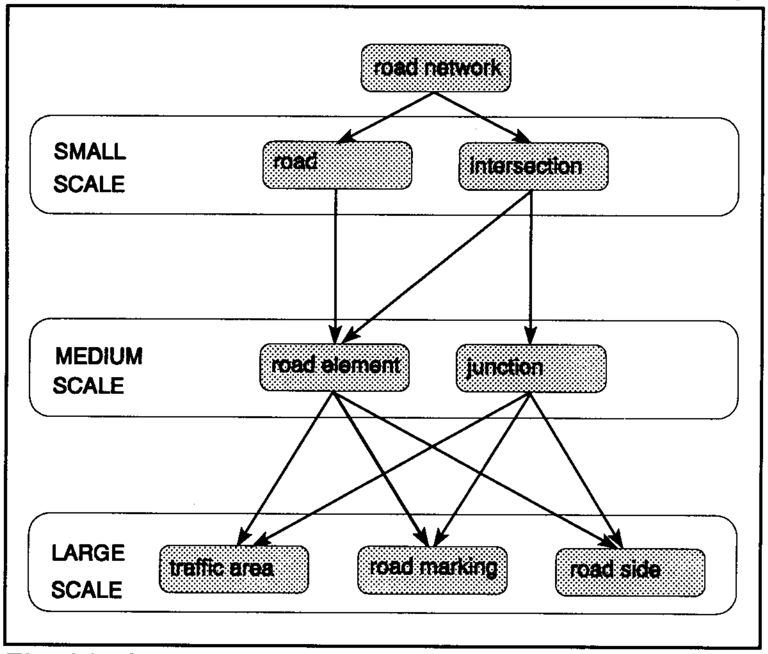

***Fig. 4.1** Object classes defined at the three levels (small, medium and large scale) of a part-whole hierarchy for road networks*

Objects at each level in their turn can be seen as generalized object classes at the top of a specialization hierarchy. For example:
- *junction* is a generalization of *crossing, fly-over* and *Y-junction*;
- a *crossing* is specialization of a *junction.*

Geometric and radiometric attributes are attached to generalized objects if they are valid for all specialized objects. Attributes attached to specialized objects serve to discriminate them from each other.

Interpretation of a particular aerial image usually produces objects belonging to the same level, which depends on the resolution of the image and the objects wished to be present in the digital map. Updating of an incomplete digital map may involve more levels if the map is acquired at a smaller scale than the aerial image. Interpretation of aerial images aims in this case to deliver a more detailed and accurate digital map, using the map at the smaller scale for prediction of parts of objects. The definitions of the object classes at each level of the part-whole hierarchy and their specialization of object classes will be given in the next sections.

4.3.1 SMALL SCALE OBJECTS

At small scale a road network consists of intersections and roads between those intersection. Intersections are considered to be points and roads are represented by lines. The physical appearance is not considered. A road consisting of physically separated carriageways is seen functionally as a single connection between two intersections. A single intersection may consist of multiple road elements and junctions at the level of medium scale. Three specialized types of roads are discriminated:
- motorways;
- main roads;
- other paved roads.

Motorways are especially constructed for fast motor traffic and have limited access from other roads by means of intersections. Main roads are defined as those roads to which traffic from motorways is able to flow to at intersections. Places where other paved roads pass over or under motorways look like an intersection in 2D, but are actually no real intersections because traffic cannot flow from one road to the other and both roads continue unchanged.

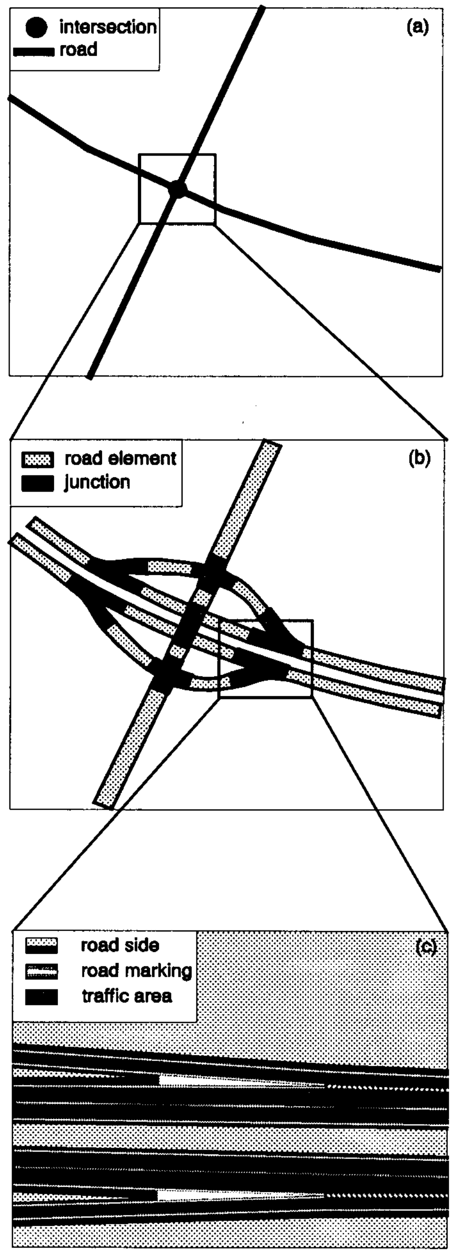

Fig. 4.2 Example of the same situation at various scales with corresponding objects (a) small scale, (b) medium scale, (c) large scale

At this level only specialized types of intersections with motorways are specified, because these become interesting in the case study (chapter 5 and 6), related to the practice of the Survey Department of Rijkswaterstaat. The next terms for types of intersections are defined in order to discriminate the following situations:
- motorway intersection: intersection between two motorways, which both continue;
- T-intersection: intersection between two motorways of which one ends at the other;
- interchange: intersection between a motorway and a continuing main road;
- T-interchange: intersection between a motorway and an ending main road;
- overpass: "intersection" between a motorway and another paved road at which traffic cannot flow from one road to the other.

There are standard solutions [Rijkswaterstaat, 1975] for the construction of intersections. A standard solution for a motorway intersection is a cloverleaf. The aerial image in fig. 5.1 depicts a standard solution for a T-intersection. Standard solutions for (T-)interchanges are a semi-cloverleaf (fig. 5.4) and a "Haarlemmermeer" solution (fig. 5.5), called after one of the places where it is applied. At so-called overpasses traffic passes over or under the motorway.

4.3.2 Medium scale objects

Intersections usually contain many road elements and road junctions to manage the traffic flow. Roads, in the sense of small scale objects, are build from road elements at medium scale (fig. 4.2b). A road element is defined as a straight or curved stretch of road without any branches. The branchpoints in the medium scale road network are called junctions. Road elements and junctions are both represented by planes. The next types of junctions are discriminated:
- crossing: the place where flows of traffic in different directions meet each other at one level and can interchange, usually guided by traffic lights;
- roundabout: traffic flows interchange by moving in one direction round a central island;
- Y-junction: the point where traffic joins or leaves the motorway by means of a link road. Considering the configuration of the first parts of the road elements at this point, it has the shape of a Y and is therefore called a Y-junction;
- fly-over: the place where a bridge carries one road over another road at a different level.

The main discrimination of road elements at this level is between main carriageways of a motorway, link roads and service roads. The main carriageway is defined as the part of the road intended for through traffic. If both carriageways are physically separated by a central reservation, each carriageway of such a so-called dual carriageway is represented by an individual road element with one-way traffic only. There are dual carriageways with 2, 3, or 4 lanes in each direction. Single main carriageways are defined as carriageways with two-way traffic. Within this group a distinction can be made between 2x2 (two lanes for each direction), 2x3 and 2x4 lanes.

Link roads are road elements which enable traffic to flow from one motorway to another, which intersects at a different level. They have one or two lanes and join or leave a main carriageway at a Y-junction. Service roads run parallel to the main carriageway and have to be used to reach houses, shops, etc.

4.3.3 LARGE SCALE OBJECTS

The road side is the imaginary line that bounds the road pavement. A traffic area is defined as a region on the road surface with a special function for traffic. The next traffic areas will be used further:
- traffic lane: elongated region with parallel sides where one is supposed to drive;
- hard shoulder: lane on which only in exceptional circumstances traffic is allowed to drive or stop;
- correction strip: narrow elongated region next to the road side meant to give drivers the opportunity to correct their course in case there is no hard shoulder;
- slip-road: elongated taper lane used to join or leave the motorway at a Y-junction.

The width of these traffic areas is defined accurately in the standards for road construction [Rijkswaterstaat, 1975], but differs depending on the fact if the traffic area is for example part of a carriageway or link road at medium scales.

Road markings are symbols painted on the road surface in order to guide, warn and control the traffic flow. There are many of them, but next (fig. 4.6) are used in the example in this chapter:
- triangular mark, which indicates that a slip-road branches off the main carriageway by means of a link road;
- edge line, a white unbroken line which marks the separation between the outermost lanes and the hard shoulder or correction strip;
- lane line, a broken line between two traffic lanes;
- block line, row of square blocks to indicated the separation between a traffic lane and a slip-road.

4.3.4 RELATIONS BETWEEN SPECIALIZED OBJECT TYPES

Spatial relationships are defined between specialized object classes, for example edge lines and lane lines can be defined to run parallel. These relationships define spatial constraints between objects on the same level, which can be used for recognition and to search one object from the other. Specialization-generalization relations define semantic constraints. Semantic constraints can be defined between objects on multiple levels and between objects on the same level.

Examples of constraints between objects on small, medium and large scale are:
- The type of intersection and solution at the small scale gives semantic clues for the detection of road elements and junctions at the medium level. For example, if an intersection is mapped at large scale as an interchange with a Haarlemmermeer solution, one should find at the medium level: four Y-junctions with link roads, two crossings and a fly-over (fig. 4.2b).
- The width of a carriageway constraints the possible number of lanes at large scale, while for example the width of the correction strip can be used to discriminate between main carriageways and link roads.

The explanations with the various types of intersections at small scale already illustrate the notion of semantic constraints at the same level. Main carriageways of motorways do not intersect each other or other roads by crossings or roundabouts, but by a fly-over, usually in combination with Y-junctions and link roads. Link roads start at a Y-junction and usually end at a level crossing or another Y-junction. Traffic areas, road markings and the road side at large scale are mutually heavily constrained. An example of a spatial constraint is the parallelism of road side, edge lines and lane lines at distances constrained by the standardized width of the traffic area in between. An example of a semantic constraint is that there is an edge line between a hard shoulder and a traffic lane.

4.4 CONCEPTS FOR MAP-GUIDED INTERPRETATION

To update an outdated or incomplete digital map three different tasks can be distinguished:
1. Change detection (section 4.4.1): Determination for every (part of the) object from the outdated map if and how it changed in the aerial image.
2. Component detection (section 4.4.2): Detection of parts from which the whole object in the incomplete map is build up.
3. Contextual reasoning (section 4.4.3): Recognition of new objects which have a spatial relationship with changed objects or with objects from the incomplete map.

How these tasks fit in the updating process of outdated and incomplete maps is the subject of section 4.4.4.

4.4.1 CHANGE DETECTION

The interpretation strategy to detect changes is based on the perception that there is a limited number of possible changes of one object type into another. For example, a carriageway with two lanes can get one or maybe two additional lanes and change into a three-lane or four-lane carriageway, but will never be reduced to a carriageway with only one lane. Hypotheses to detect changes are initiated by objects present in the outdated map.

4.4.2 COMPONENT DETECTION

Component detection assumes that the incomplete map is acquired on a smaller scale at which it is not possible to map parts from which the whole objects are build. The current goal is to detect in an aerial image of larger scale those objects which have part-whole relationships with the mapped objects. Usually initial hypothesis generation based on part-whole relationships only aims to detect some of the components with strong semantic constraints. The rest is searched by contextual reasoning, in this way also including spatial constraints.

4.4.3 Contextual reasoning

The strategy of contextual reasoning is based on the perception that each object type is spatially related to one or more other object types. For example, a road marking like an edge line has a certain distance to another parallel edge line and eventually to one or more parallel lane lines. These relations are used to generate hypotheses for top-down search of objects. Detection of one of these new objects immediately generates hypotheses for the presence of other spatially related objects. But also detected changed objects or objects from an incomplete map may generate hypotheses for spatially related objects. Thus a chain of search actions emerges. The mechanism to handle these search actions will be discussed in 4.5.1.

Spatial relationships are not only used for top-down hypothesis generation, but also for bottom-up consistency verification. Verification is based on the fact that some spatial relationships yield conditions between two objects. If the first object is found, the second object **must** be found or is prohibited to appear in its vicinity. An example of such a condition is that the detected change of a part of a motorway into a Y-junction needs to be confirmed by the presence of a new link road. This case will be discussed extensively in chapter 5 and 6. If this condition is not fulfilled, even after reprocessing, the classification of the first object is rejected.

4.4.4 Map-guided interpretation strategy

The interpretation strategy defines which tasks are performed and in which order. A requirement for the interpretation strategy designed in this thesis is that the digital maps which are updated, also guide the interpretation process. The interpretation strategy differs for outdated and incomplete maps. Change detection and contextual reasoning are the tasks performed in the interpretation strategy for updating of outdated maps, while in case of updating incomplete maps component detection and contextual reasoning are used. The interpretation strategy for both cases will be worked out in more detail.

In case outdated maps are updated, processing starts with change detection. Change detection includes generation of hypotheses for possible changes, depending on which objects are present in the map and verification of these hypotheses by starting top-down processes for segmentation of the aerial image. Thus objects from outdated maps and the aerial image are input for change detection. This results in either changed or unchanged objects. Unchanged objects will be part of the new map. If changed objects are not spatially related to other objects, then the interpretation process stops and changed objects are appended to the new map. An example of this case is that the addition of an extra lane to a road without branches has no consequences for other roads. This case will be shown in section 4.8.4 as an example of change detection only.

If, on the other hand, changed objects are spatially related to other objects, they are input for the process of contextual reasoning to search spatially related objects, together with the aerial image. This case is shown in fig. 4.3. Contextual reasoning results either in detection of a new object or not. If after several adaptations there is no object detected, but there is a constraint for the presence of the new object in relation to the changed object, the hypothesis for a change is still rejected. This results in still appending the unchanged object to the new map. Each detected new object is appended to the new map, but also restarts the cycle of contextual reasoning until all related objects are searched for. This case is the subject of the case study in chapter 5 and 6.

In case incomplete maps need to be updated, objects from this outdated map and the aerial image are input for component detection, followed by contextual reasoning. An example of this case is worked out in section 4.8.3.

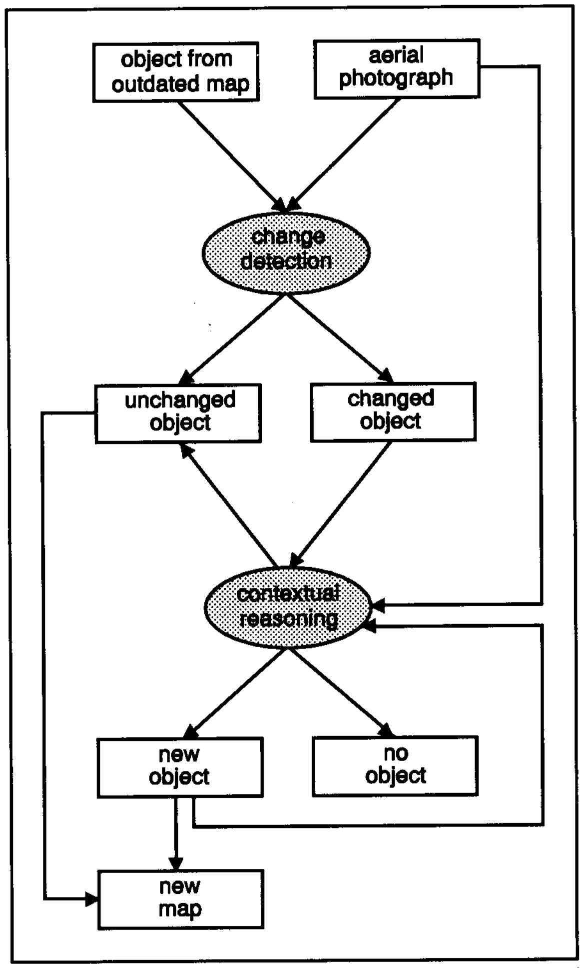

Fig. 4.3 Process for updating of an outdated digital map using an aerial image.

4.5 REALISATION OF THE INTERPRETATION STRATEGY

The interpretation strategy is based on the classic hypothesize and test paradigm. Within this strategy top-down and bottom-up processes are integrated. The control cycle has four processing steps (fig. 4.4):
- (alternative) hypothesis generation;
- goal-directed segmentation;
- object recognition;
- inconsistency detection.

The requirements, which are described in section 4.1, are incorporated in this strategy. The next sections will outline these processing steps of the designed interpretation strategy in more detail.

This implementation of concepts was primarily influenced by some of the design criteria of the framework for map interpretation by Hartog [1995]. Maps can be seen as a simplified representation of the real world depicted in the aerial photograph. Therefore it is possible to use the same control strategy as for searching objects in scanned maps, extended with more sophisticated segmentation algorithms and specific knowledge about road networks. In order to update outdated digital maps the interpretation strategy needs to be modified such that not only segmented objects, but also objects from the outdated map can be input of the interpretation strategy.

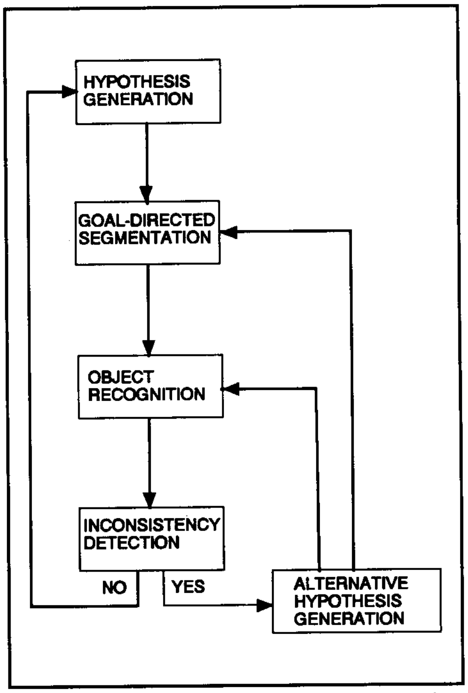

***Fig. 4.4** Control cycle with processing steps*

4.5.1 HYPOTHESIS GENERATION

In the context of map-guided interpretation, hypotheses concern changes or new objects which are possibly present. Reason for their hypothesized presence is their relationship with objects in the map or with objects detected in the previous processing step. For example, when the road side is detected, the presence of road markings like edge or lane lines parallel to the road side can be hypothesized. All possible relationships are defined in the frames. An example of a small semantic network underlying the frames for searching road markings on a Y-junction is shown in fig. 4.5.

A task during hypothesis generation is to determine which new hypotheses are relevant due to the last processing step and in which order they should be verified, also considering previously generated hypotheses. Realization of the hypothesis generation mechanism is done by means of a search list, which can be regarded as an ordered list of processes to be activated. Search actions to all objects, associated with just detected objects or objects in the map by an object relation, are appended to the list. Actions in the list are ordered by their priorities, associated with each object relation (see fig. 4.5). These priorities are assigned a priori to the object relations when constructing the knowledge base. Processing continues until all hypotheses in the list are verified and the search list is empty.

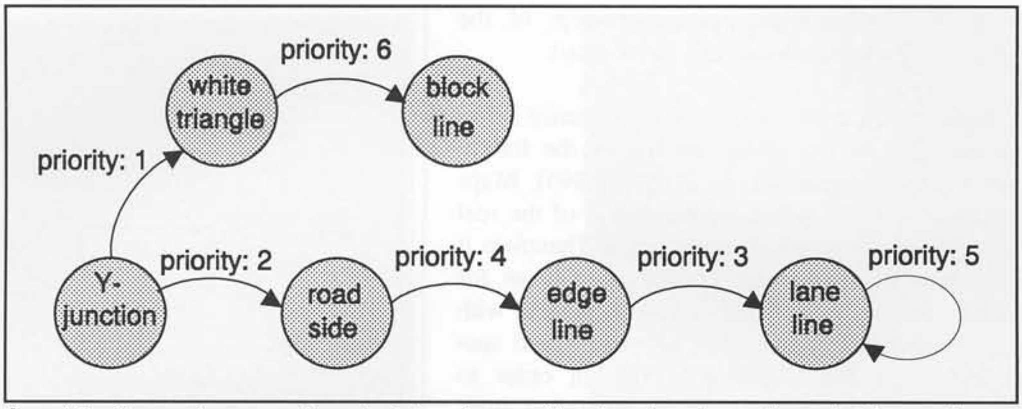

fig. 4.5 Semantic network underlying frames for the detection of road sides and road markings from a medium scale Y-junction in an incomplete map. Relations correspond to hypotheses which can be generated.

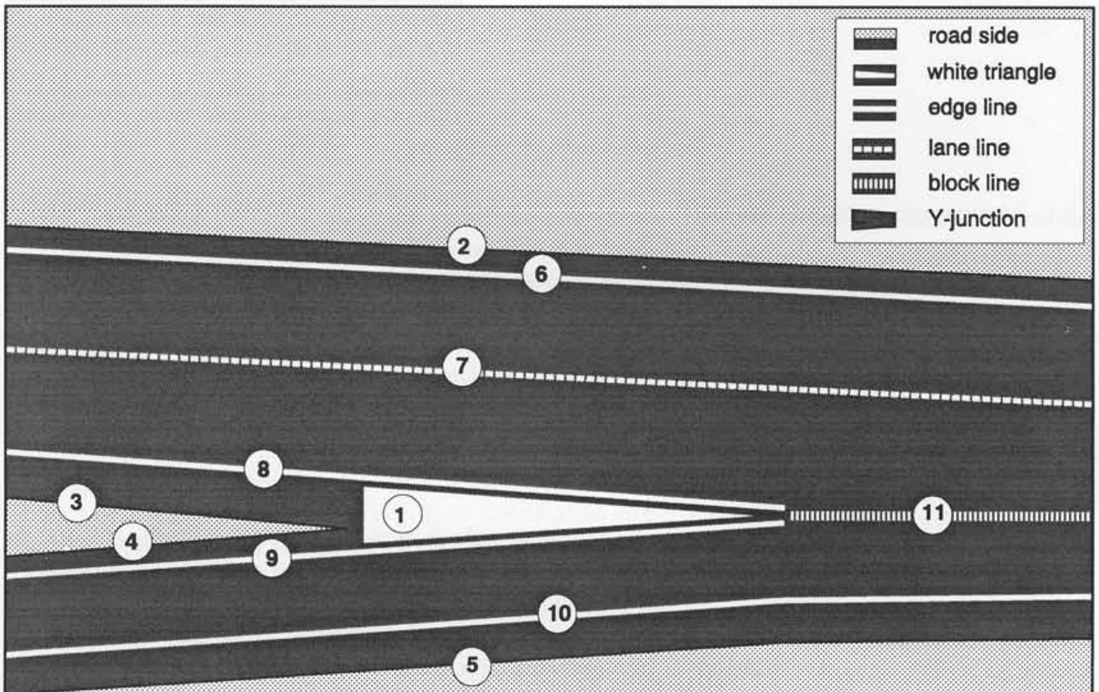

Fig. 4.6 *Artificial large scale photograph of a Y-junction. Numbers relate to the sequence in which road sides and road markings are found according to the semantic network in fig. 4.5.*

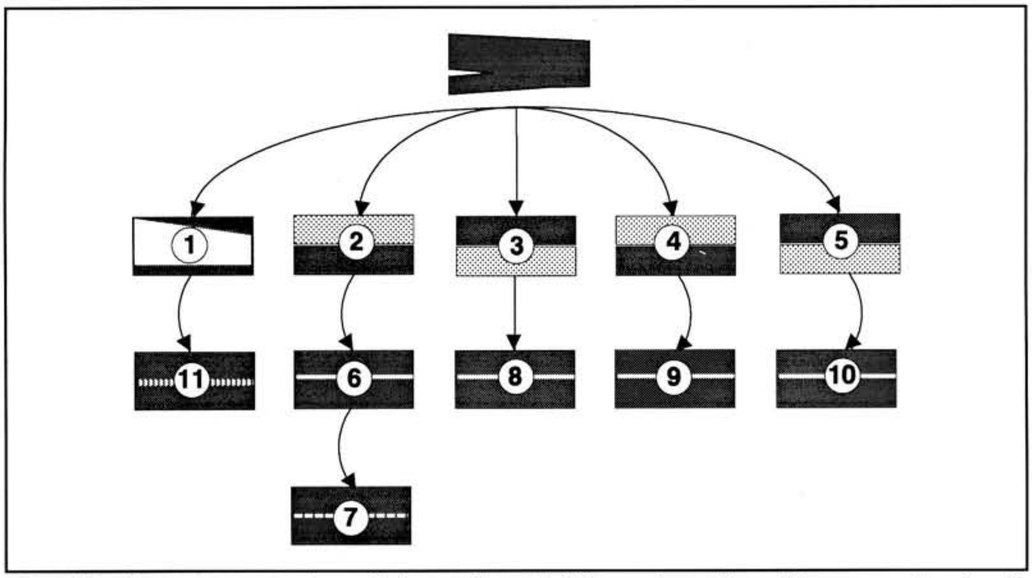

Fig. 4.7 Tree representation of detected road sides and road markings associated with the objects it caused and was caused by. The numbers correspond to fig. 4.6.

The process of hypothesis generation will be illustrated with an example. Consider the artificial large-scale photograph of a Y-junction in fig. 4.6. Using the Y-junction originating from medium scale, the aim is to complete this map with road sides and road markings extracted from a large scale image. The semantic network in fig. 4.5 represents the relationships used for generation of hypotheses. The hypothesis generated by the relationship with the highest priority, corresponding to the lowest value, is put on top of the search list and will be activated first. Two relationships are associated with the Y-junction: a triangular mark (priority 1) and the road side (priority 2). This results in two generated hypotheses which are put in the action list. The search action for a triangular mark is on top, since it has a higher priority. Assume that activation of this hypothesis leads to detection of a triangular mark, number 1 in fig. 4.6. Next, the hypothesis for detection of a block line is generated. It has a low priority (6) and is put at the bottom of the list. Consequently, the action to search the road side is now activated first. It results in the detection of four pieces of road side, number 2, 3, 4 and 5. Each of them generates a hypothesis for the search of an edge line. Because all four hypotheses have the same priority (4), their sequence in the search list is arbitrary, but they will all be put above the search action for the block line (priority 6). Assume that action associated with road side number 2 will be on top of the list. It results in the detection of edge line number 6. Next the hypothesis for a lane line is generated. Its priority (3) is higher than all the other actions in the list and it will be activated first. The detection of a lane line (number 7) generates hypotheses for other lane lines, which are put on top of the list. However, no more lane lines are found. The remaining actions for the search of edge lines are successively activated. Detection of each edge line number 8, 9 and 10 also results in search actions for lane lines. Since no further lane lines are present, they will not be found. Finally, the action for search of a block line is on top of the list. Although it results in detection of a block line (number 11), no more hypotheses are appended to the action list because there is no relationship represented in the semantic network, starting from a block line. Thus, the action list remains empty and processing stops. Another representation of this process is given in fig. 4.7, where every object is connected to the one it is detected from and the one it detects itself.

4.5.2 Segmentation

Activation of generated hypotheses means goal-directed segmentation of the hypothesized object in the aerial image. This requires segmentation techniques tailored specially for the optimal detection of a particular object type. Their a priori specification, together with their parameters and a search area, are associated with relations in the knowledge base between objects. A specific goal-directed segmentation technique is only activated if the hypothesized object is searched from the object with which the technique is associated in the knowledge base. Because segmentation is never perfect, it is possible to define alternative parameters, alternative segmentation techniques and an alternative search area. Hypotheses for re-segmentation are generated if an inconsistency is detected.

4.5.3 Object recognition

After top-down segmentation, properties of the segmented object are compared with a priori knowledge in a bottom-up process. Values of geometric and radiometric properties should be within the range of values of the corresponding attributes of the a priori defined object. Properties for object recognition can be defined in the object definitions as well as in the object relation (see section 4.6). If properties are related to the segmentation technique, they are usually defined in the object relation. Properties generally valuable for a certain object are usually defined in the object definition.

4.5.4 Inconsistency detection

For realisation of inconsistency detection distinction is made between optional and essential relations. An optional relationship between two objects indicates that if the first object is found, it is likely to find the second in its vicinity. An essential relationship implies that if the first object is found, the related object *must* be found also or else the first object is wrong. Essential relations provide a mechanism to detect inconsistencies in the interpretation. These inconsistencies are handled by generation of alternative hypotheses, resulting in re-segmentation.

4.6 KNOWLEDGE REPRESENTATION

4.6.1 BASIC REPRESENTATION PRIMITIVES

The represented knowledge is divided into two different types [Hartog, 1995]: object definition and object relation.

The *object definitions* contain a description of geometric and radiometric properties for each object type. They are used to test hypotheses of goal-directed segmentation.
The *object relations* describe spatial relationships between object types and changes of one object type into another, in this way constructing a complex semantic network which describes both declarative and procedural knowledge. The declaration part describes the location of objects with respect to each other, while the procedural part of the knowledge specifies where to search and with which image processing technique. Object relations are used to control change detection and contextual reasoning.

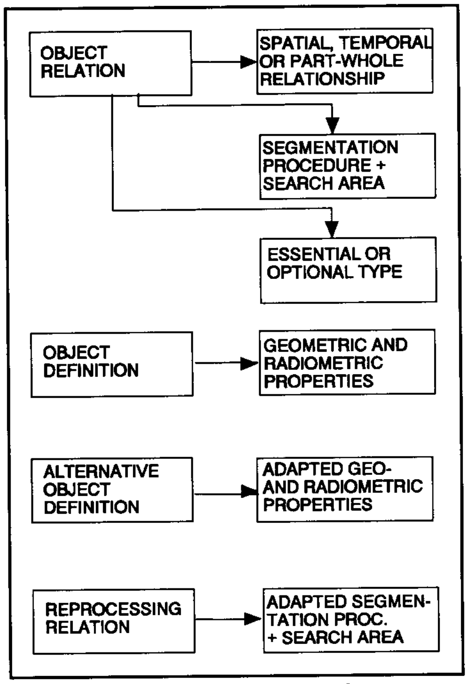

Fig. 4.8 *Types of knowledge represented in object definitions and object relations*

Variations upon object definitions and relations, which are used for generation of alternative hypotheses in case of inconsistency detection, are called alternative object definitions and reprocessing relations respectively. The contents of object definitions and object relations will be discussed in more detail in the next sections. An overview of the types of knowledge they contain is given in fig. 4.8.

4.6.2 OBJECT DEFINITION

Object definitions contain a priori knowledge represented in frames. An object definition consists of a name, an origin, and a set of numerical features. The origin discriminates objects of the map from objects extracted by segmentation. In this way it is possible to search an object from the map in the image by defining a relation between the object from the map and the object in the segmented image. A standard set of geometric features is available [Hartog, 1995], but it is also possible to define other features, like features based on radiometric properties. These properties should be chosen such that they depend as little as possible on a specific test image. Only relevant properties are specified. A range of allowed values is attached to each feature, providing a condition for recognition of the object. Because this range should be defined a priori, extensive testing will be necessary to determine appropriate values. An example of an object definition is given in fig. 4.9.

```
DEFINE OBJECT
    name: triangular mark
    origin: segmentation
// geometric conditions:
    area/(length×height)      [0.40, 0.60]
// radiometric conditions:
    average grey value:       [≥ 200]
ENDDEF
```

Fig. 4.9 Object definition triangular mark

```
DEFINE ALTERNATIVE
    name: triangular mark
    origin: segmentation
// geometric conditions:
    area/(length×height)      [0.35, 0.65]
    length                    [> 40 m.]
// radiometric conditions:
    average grey value:       [≥ 200]
ENDDEF
```

Fig. 4.10 Alternative object definition for triangular mark

This object definition contains one geometric and one radiometric property. Since road markings in the Netherlands are always painted in white, the average grey value of the segmented object will always be high in intensity images. In theory the area of the triangle divided by its height times its base length is exactly 0.5. However, because of the discrete character of pixels, some deviation should be allowed in this condition. Based on experience a deviation of 0.10 was defined. For an acute-angled triangle the deviation due to discretisation may even be larger (0.15). Therefore an alternative object definition can be used (fig. 4.10).

4.6.3 OBJECT RELATION

Each object relation contains four parts:

- an identification part which specifies the type of relationship and between which objects the relation is defined;
- a part for goal generation containing the priority of the relationship and characteristics of the relationship (optional or essential) for inconsistency detection;
- a part for goal-directed segmentation which specifies an image processing technique, its parameters and a search area;
- a part for object recognition, containing geometric and radiometric features of the relationship or features associated with the specified image processing technique.

In the context of map updating there are three types of relationships:

```
DEFINE RELATION
    type:   changed into
    from:   two lane motorway
    into:   three lane motorway

// part for goal generation
    priority:   1
    optional: YES

// part for goal-directed segmentation:
    image processing technique:
    - SearchExtraLane, about x m. width,
    in y direction
    parameters: x = 3.0 m., y = vertical
    search area: parallel to the road
    element, width = 3.5 m.

// part for object recognition
// spatial relationship
    100% overlap
ENDDEF
```

Fig. 4.11 Definition of temporal relationship

- "temporal" (fig. 4.11), between an object in the outdated map and an object in the image into which it possibly changed;
- "spatial" (fig. 4.12), between two objects in the image, which have a contextual relationship.
- "part-whole" (fig. 4.13), between an object from an incomplete map and an object in an image which shows more details.

Either standard techniques for image processing (fig. 4.12 and 4.13) or a technique specially developed for the goal specified in the relationship (fig. 4.11) can be used.

```
DEFINE RELATION
    type:   spatially related to
    from:   road side
    to:     edge line

// part for hypotheses generation
    priority:   2
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - maximum cost path calculation by
    dynamic programming
    parameters: 15 pixels used to calculate
    local direction of the road side
    search area: parallel to road side at
    distance of 3.0 m.

// part for object recognition:
// geometric conditions:
    more vectors than road side [< 10 ti-
    mes]
ENDDEF
```

Fig. 4.12 Definition of spatial relationship

```
DEFINE RELATION
    type:   consists of part
    from:   Y-junction
    to:     triangular mark

// part for goal generation
    priority:   1
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - thresholding grey values > x
    - median filtering
    - remove objects < y pixels
    parameters: x = 200, y = 500
    search area: inside Y-junction

// part for object recognition:
    empty
ENDDEF
```

Fig. 4.13 Definition of part-whole relationship

Reprocessing relations define alternatives for essential object relations, which result in re-segmentation if the object is not detected. Therefore they are associated with an object relation. As a consequence they do not contain the part for goal generation and object recognition, but only the part for goal-directed segmentation. Fig. 4.14 shows a reprocessing relation of 4.12 which defines an adapted search area. The example in section 4.8.2 illustrates the use of both relationships to search an edge line from the road side in an aerial image. Fig. 4.15 shows an alternative of relation 4.13 with adapted parameters. The example in section 4.8.3 uses these relationships to illustrate the principle of alternative parameters and segmentation techniques on another aerial image.

```
DEFINE REPROCESSING
    type:   spatially related to
    from:   road side
    to:     edge line

// part for goal-directed segmentation:
    image processing techniques:
    - maximum cost path by dynamic pro-
    gramming
    parameters: 15 pixels used to calculate
    local direction of the road side
    search area: parallel to road side at
    distance 1.1 m.
ENDDEF
```

Fig. 4.14 *Definition of reprocessing relation with alternative search area for relation 4.12*

```
DEFINE REPROCESSING
    type:   consists of part
    from:   Y-junction
    to:     triangular mark

// part for goal-directed segmentation:
    image processing techniques:
    - thresholding grey values > x
    - median filtering
    - remove objects < y pixels
    parameters: x = 220, y = 500
    search area: inside Y-junction
ENDDEF
```

Fig. 4.15 *Definition of reprocessing relation with alternative parameters for relation 4.13*

4.6.4 REPRESENTATION OF SEGMENTED OBJECTS

Objects extracted from the image should be stored in such a way that their properties can be easily compared with the conditions for recognition defined in the object definitions, but also such that raster information can easily be retrieved for example for definition of the search area for low level processing. These instantiations of objects are only stored temporarily and are therefore called STM objects. Also objects from the map are in advance converted to STM objects, because they need to be compared with objects in the knowledge base to start hypothesis generation.

STM objects are stored in a linked list, which allows fast searching for objects with specific characteristics, which Nagao [1984] considered to be a problem. Each STM object has six groups of attributes (fig. 4.16):
1. identification: concerns a unique label number which is assigned;
2. classification: indicates of which defined object type the STM object is an instance after it is recognized and what its origin is (map or segmentation);
3. iconic description: a raster image of the segmented object with the size of its minimum bounding rectangle together with its location in the image;
4. symbolic description: geometric and radiometric attributes, which can be used for recognition;
5. vector description: coordinate descriptions of the contour and axis of the object, if they were calculated during segmentation;
6. related objects: lists of label numbers of objects with a special relationship with the concerned STM object, for example touching, near, part_of, resulted_in or caused_by.

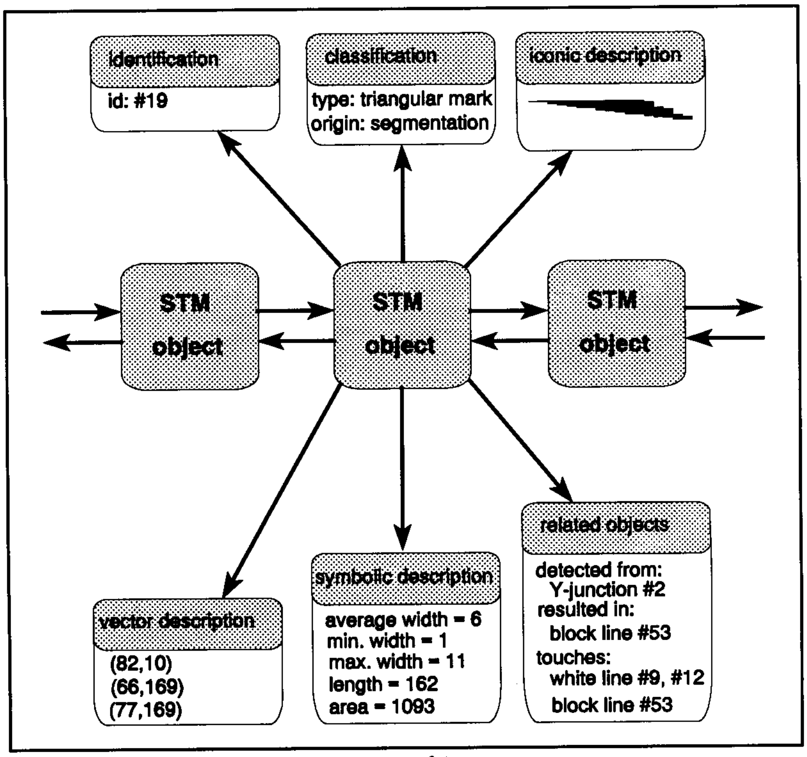

Fig. 4.16 Representation of a STM object

4.7 COMPLETE INTERPRETATION PROCESS

A scheme for the complete interpretation process with associated knowledge sources, input and output data is shown in fig. 4.17. It is an integration of fig. 4.4 and 4.8. It relates particular types of knowledge and input/output data to certain tasks in the interpretation strategy.

Hypothesis generation uses initially only objects from the outdated road map for which a temporal relationship is established as object relation in the knowledge base or objects from the incomplete map included in a part-whole relationship. Their object relations define which image processing routines will be used for goal-directed segmentation of the aerial image, their parameters and search area. Geometric and radiometric properties defined in the knowledge based are compared with corresponding features of the STM objects created after segmentation. If a priori defined properties correspond with properties of the segmented object, the object of searched type is found and appended to the updated road map. If the object is not found, an inconsistency is detected if the object relation is essential. In this case either an alternative object definition is used for object recognition, or another search area and/or other (parameters of the) image processing techniques are used for re-segmentation.

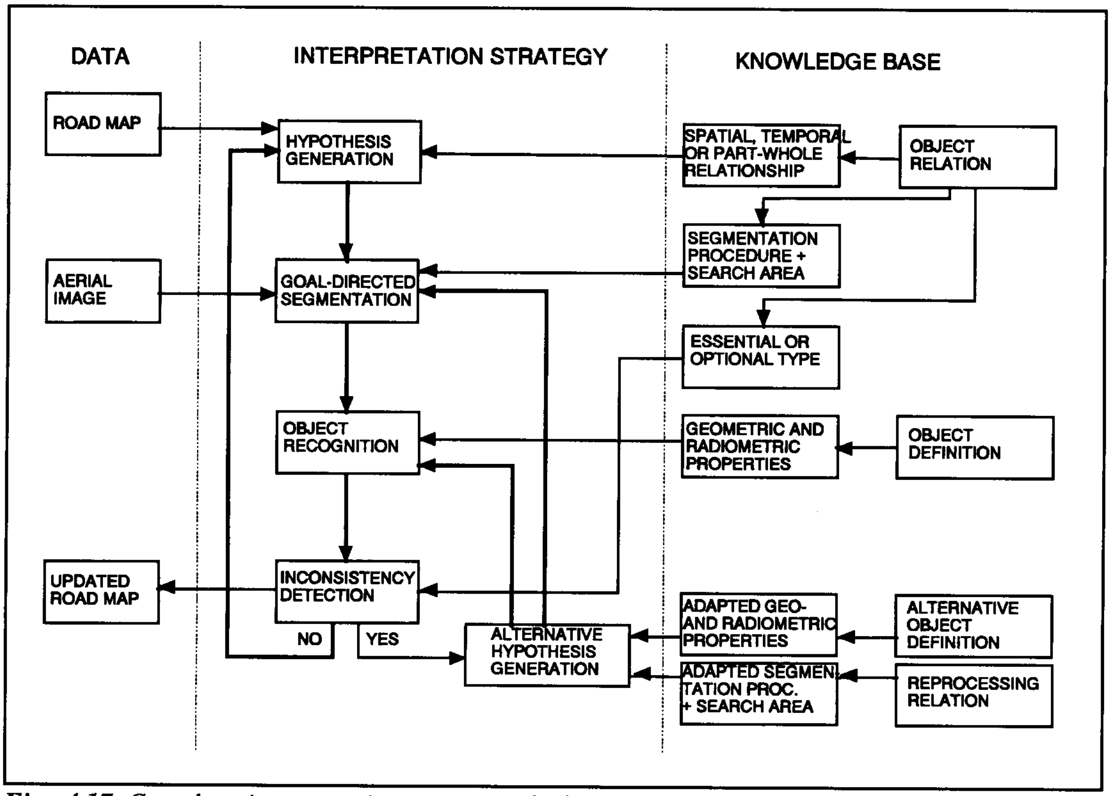

Fig. 4.17 Complete interpretation process (thick arrows) with associated knowledge sources, input and output data (thin arrows)

4.8 EXAMPLES

This section presents examples of the application of the interpretation strategy on aerial images. The contents of the knowledge bases used in these examples mainly consists of the object definitions and relations used to illustrate the concepts in the previous sections. The first three examples deal with contextual reasoning at large scale. The examples in section 4.8.1 and 4.8.2 aim to illustrate the use of alternative object definitions and reprocessing relationships. The third example shows a complete interpretation of all road markings present in an aerial image using their spatial relationships. The example in section 4.8.4 deals with change detection at medium scale, not followed by contextual reasoning. Integration of both change detection and contextual reasoning will be the subject of the case study in chapter 5 and 6.

Ranges of attributes in object definitions and parameters for goal-directed segmentation used in these examples were not tested exhaustively, because the aim of these examples is only to illustrate the concepts.

4.8.1 EXAMPLE OF ALTERNATIVE HYPOTHESES GENERATION

To illustrate the concept of alternative object definitions, alternative parameters and segmentation techniques for goal directed segmentation, the next example is used: searching a triangular mark from a medium scale Y-junction. This relationship is part of the semantic network in fig. 4.5. The frame associated with this relationship was already defined in fig. 4.13. Since road markings in the Netherlands are always painted in white, they can be extracted from an intensity image by thresholding pixels with a high grey value. The value 200 was chosen as threshold. Next a median filter, size 3x3, is applied to remove small holes and protrusions due to near road markings. Finally small objects are removed. Fig. 4.15 already showed a relationship with adapted parameters for re-segmentation, using the value of 220 as threshold. Another relationship, which defines a different image processing technique for re-segmentation, is shown in fig. 4.16. Thresholding a few percent of pixels with the highest grey values is less sensitive for absolute grey values. From the standards for road construction [Rijkswaterstaat, 1975] it can be calculated that linear road markings cover 2-3% of the road surface. A threshold of 2.5% turns out to work quite well.

```
DEFINE REPROCESSING
   type:   consists of part
   from:   Y-junction
   to:     triangular mark

// part for goal-directed segmentation:
   image processing techniques:
   - thresholding highest x%
   - median filtering
   - remove objects < y pixels
   parameters: x = 2.5%, y = 500
   search area: inside Y-junction
ENDDEF
```

Fig. 4.18 Relation for re-segmentation with a different image processing technique

	STM object number			
	1	2	3	4
Area (in m²)	181	259	196	171
Length (in m.)	64	88	77	76
Width (in m.)	4.7	4.4	4.0	4.0
=> $\frac{Area}{Length \times Width}$	0.596	0.673	0.635	0.568
Average grey value	244	239	246	249

***Table 4.1** Relevant properties of resulting STM objects using alternative hypotheses*

Fig. 4.19a shows a part of an aerial image, ground resolution 0.4 m., which contains two Y-junctions. Activation of the hypothesis defined by the relation in fig. 4.13 results in two STM objects (1 and 2) drawn in black in fig. 4.19b and projected on top of the Y-junctions from a digital map. These STM objects need to fulfil the conditions defined for a triangular mark in fig. 4.9. As can been seen in table 4.1 this condition is fulfilled for STM object 1, but not for 2.

Probably because the white truck which is merged with the segmented triangle. Hence STM object 1 is recognized as white triangular mark, but for STM object 2 re-segmentation with an adapted threshold value of 220, according to the relation in fig. 4.15, is done. Fig. 4.19c shows the result. STM object 3 is still not recognized as triangular mark, unless the alternative object definition of fig. 4.10 is used. Hypothesis generation based on the relation in fig. 4.18, which defines a different segmentation technique would already lead to recognition (STM object 4) with the original object definition (fig. 4.19d). The radiometric condition, an average grey value ≥ 200, is fulfilled for all STM objects.

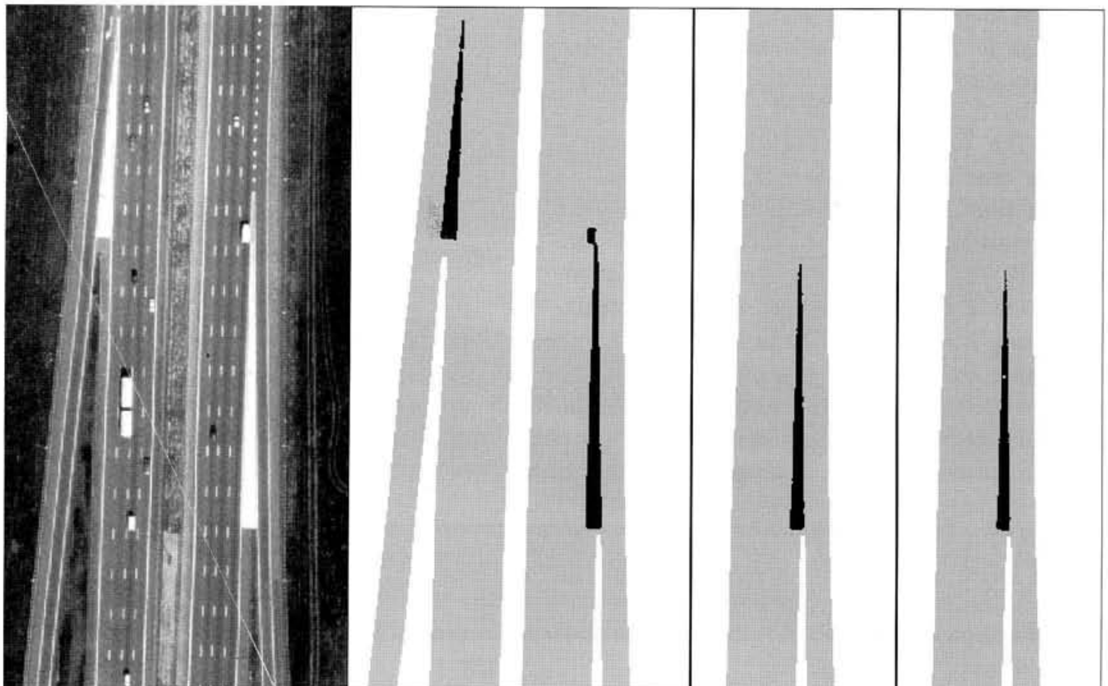

Fig. 4.19 *(a) Aerial image with 2 Y-junctions, (b) Result of thresholding with grey value 200, (c) Idem with value 220, (d) Result of thresholding highest 2.5%*

4.8.2 Example of an alternative search area

In order to illustrate adaptation of the search area, another example is used: search of an edge line from the road sides of a dual carriageway of a motorway with two lanes each. The distance between the road side and this line differs for the left and the right side of the road (right in the sense of the side one drives at in the Netherlands). There is a hard shoulder of 3.0 m. wide at the right side of the road, while at the left side there is only a narrow correction strip of 1.1 m. wide. As a consequence two different search areas need to be defined for searching a edge line: one parallel to the road side at a distance of 3.0, the other at a distance of 1.1 m. The corresponding semantic network is part of fig. 4.5 and the corresponding frames were already shown in fig. 4.12 and 4.14.

```
DEFINE OBJECT
   name: edge line
   origin: segmentation
// geometric conditions:
   thickness  [0.05 m., 0.45 m.]
// radiometric conditions:
   average intensity:        [≥ 200]
   % pixel ≥ 200  [≥ 70%]
ENDDEF
```

***Fig. 4.20** Object definition of edge lines*

In this example a dynamic programming algorithm [Gerbrands, 1988] is used, which detects the maximum cost path in a region of interest. A characteristic of this algorithm is that the path is often erratic. The geometric property defined in the object relation in fig. 4.12 is related to dynamic programming. It defines that the number of vectors after vectorization of the path should not be more than 10 times larger than the number of vectors of the road side. Since they are parallel, a much larger number of vectors indicates a very erratic path, which should be rejected. The property defining that grey values along the path should be relatively high is general valuable for edge lines, which are painted in white. These kind of properties should be defined in the object definition (fig. 4.20).

Another characteristic of dynamic programming is that always a path is found. Therefore during object recognition it needs to be determined which of the two lines found at each side of the road at a distance of 3.0 m. will be accepted as edge line. Actually the geometric condition that the thickness of the edge line should be between 0.05 and 0.45 is not relevant in case of dynamic programming, since this technique always produces a line of one pixel thick, in this case 0.40 m corresponding to the resolution of the aerial image. In general it is a very valuable property because in the standards for road construction [Rijkswaterstaat, 1975] an edge line is defined to be 0.15 m. thick. For that reason it is added in the general valid object definition.

The result of processing the part of the aerial image in fig. 4.21a is shown in fig. 4.21b and 4.21c. Searching an edge line from the inner road side at a distance of 3.0 m. resulted in the detection of the lines in fig. 4.21b, corresponding to STM object 1 and 2 in table 4.2. For both the radiometric conditions are not fulfilled. At the inner road side of a 2-lane motorway the distance between road side and edge line is actually 1.1 m. according to the standards for road construction [Rijkswaterstaat, 1975]. Consequently this hypothesis was correctly rejected.

	STM object number					
	1	2	3	4	5	6
Times more vectors	7	8	2	6	5	1
Average grey value	170	180	247	244	235	218
% of grey values > 200	10%	18%	96%	93%	90%	72%

***Table 4.2** Relevant properties of STM objects, which result from searching triangular marks*

Re-processing with a search area at a distance of 1.1 m. resulted in STM object 3 and 4 (fig. 4.21c), which were both accepted. STM objects 5 and 6 resulting from searching an edge line from the outer road side at a distance of 3.0 m. were accepted as well.

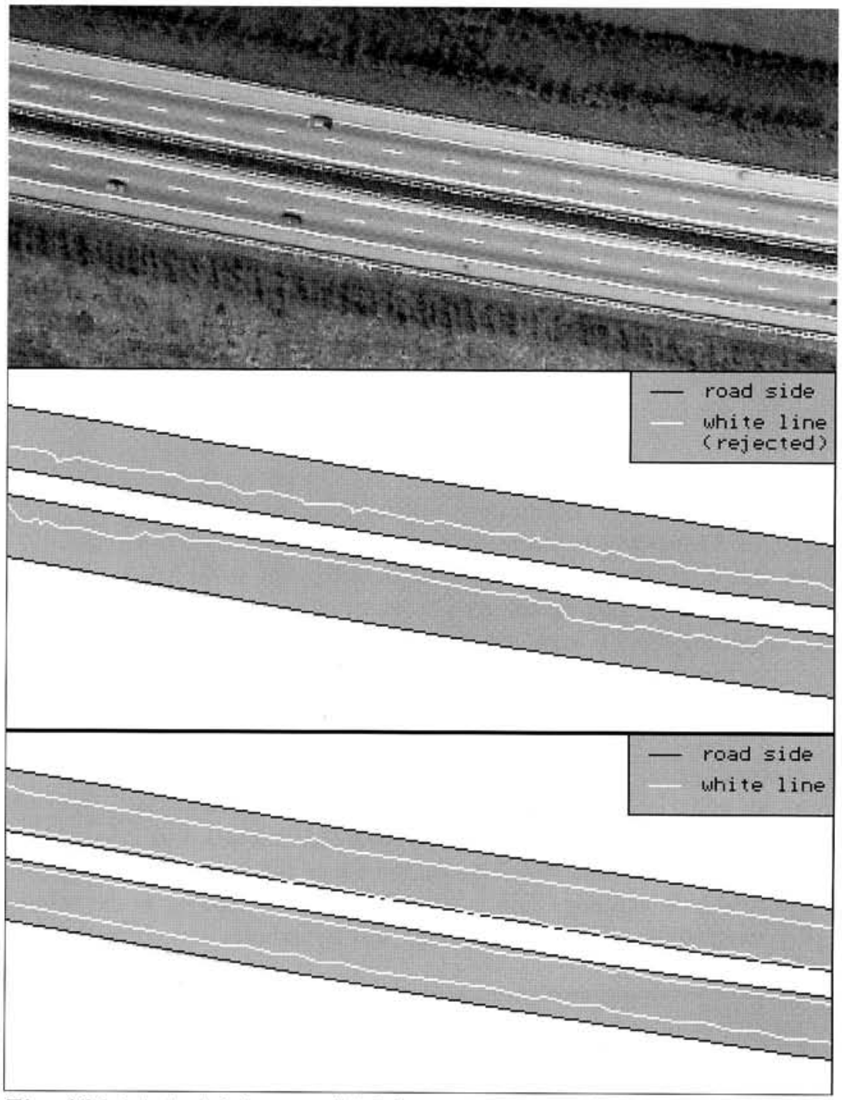

Fig. 4.21 *(a) Aerial image with 2-lane motorway, (b) Rejected edge lines detected from inner road sides, (c) Accepted edge lines at 3.0 and 1.1 m.*

***Fig. 4.22** Results at large scale of detection of road sides as components of a Y-junction at medium scale and detection of road markings by contextual reasoning*

LEGEND	
Road side	Edge line
Lane line	White triangle
	Block line

4.8.3 EXAMPLE OF COMPONENT DETECTION AND CONTEXTUAL REASONING

In this example the result of contextual reasoning is shown using the complete semantic network in fig. 4.5. Actually it is a combination of the examples in section 4.8.1 and 4.8.2.
The same aerial photograph of the Y-junction (fig. 4.19a) was used as in section 4.8.1. Results are projected in colours on the aerial image in fig. 4.22. The medium scale digital map of the Y-junction (grey in fig. 4.19b) is used to detect triangular marks (blue in fig. 4.22) and road sides (red). The road sides initiate the detection of edge lines (green), which on their turn lead to detection of lane lines (yellow). Finally, from one of the triangular marks a block line (pink) is found, since the other one is outside the image.

4.8.4 EXAMPLE OF CHANGE DETECTION

This example illustrates detection of a change at medium scale, which has no consequences for the presence of other objects. The aim is to determine if a carriageway of a motorway with two lanes changed into a carriageway with three lanes. There is only one object relation needed, which was already given in fig. 4.11. The image processing routine, SearchExtraLane, is based on area-based matching of an iteratively widening raster representation of the 2-lane carriageway from the map. As criterium for comparison of matches an increase in the cross-correlation is used. The object definitions for a 2-lane and 3-lane carriageway are respectively given in fig. 4.23 and 4.24. The geometric condition, width of the carriageway, is calculated from the standards for road construction. Another way to detect this change, which is not considered here, is to search traffic areas at large scale and use them to determine the specialized type of road element at medium scale.

```
DEFINE OBJECT
    name: two lane carriageway
    origin: map
// geometric conditions:
    width: 11.2 m.
```

Fig. 4.23 Object definition for motorway with 2 lanes

```
DEFINE OBJECT
    name: three lane carriageway
    origin: segmentation
// geometric conditions:
    width: 15.6 m.
```

Fig. 4.24 Object definition for motorway with 3 lanes

Fig. 4.25a shows an aerial photograph of a carriageway with three lanes. The outdated map, which contains a carriageway with two lanes, is projected on it in white. Applying the relation in fig. 4.11 results in detection of a 3-lane carriageway, projected in white in fig. 4.25b. In case the same relation is applied to an up-to-date map of 2-lane carriageways, projected in black on the aerial image in fig. 4.25c, the change is rejected.

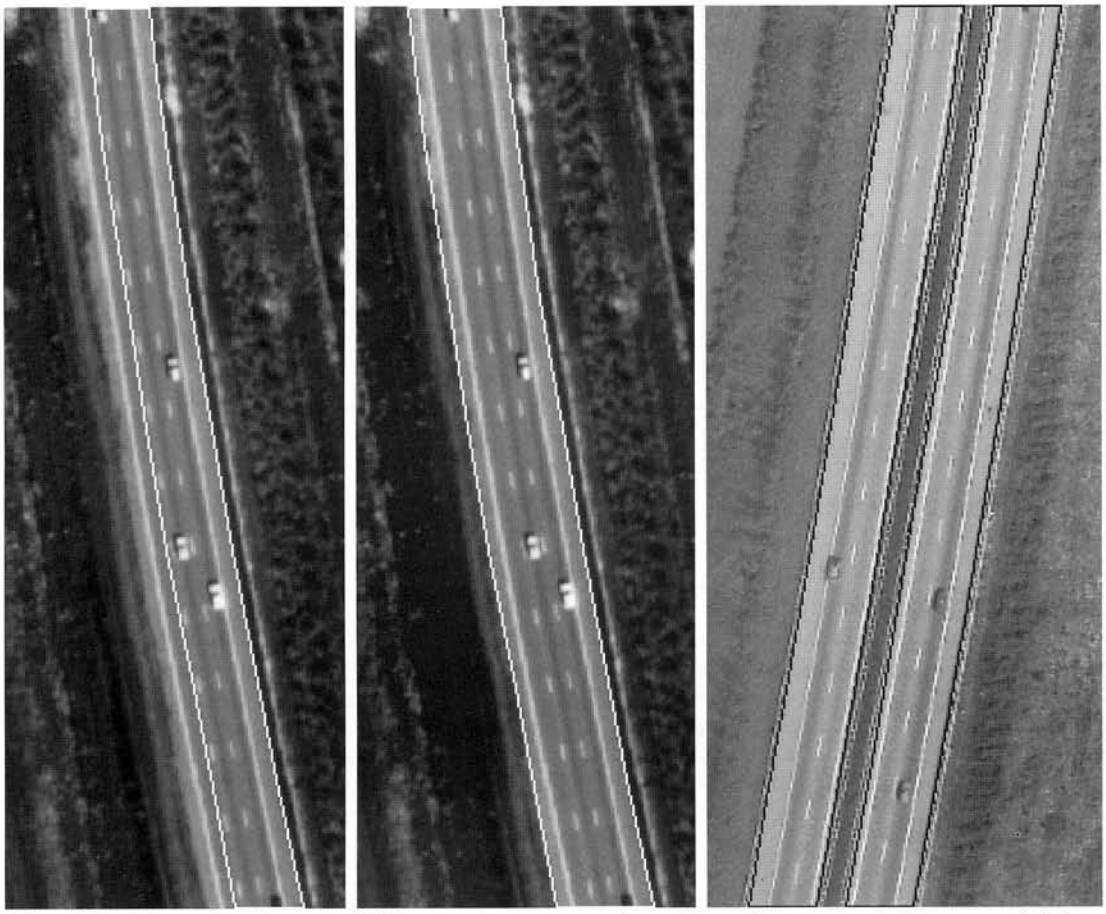

Fig. 4.25 (a) outdated map of 2-lane motorway projected in white on aerial image of 3-lane motorway, (b) result of change detection, (c) result of change detection in case an up-to-date map (in black) is used.

4.9 DISCUSSION

In general there are two possible errors in interpretation: missing object or misclassification. The definition of essential and optional relationships is used as a mechanism to detect these types of errors. If an essential relationship is not present, the missing object is searched using alternative processing or object definitions. If it is still not found, the object which initiates the hypothesis is supposed to be misclassified and its classification is rejected. To check if objects related by an optional relationship are eventually missing, extra relations can be appended. Fig. 4.26 shows a part of the semantic network of fig. 4.5 with an extra relationship (with priority 6) between a lane and an edge line for the detection of eventually missing edge lines.

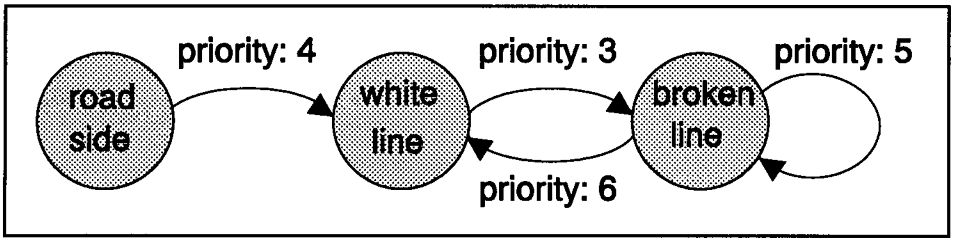

Fig. 4.26 Semantic network with extra relationship to detect missing edge lines

There is no general procedure for adaptation of image processing technique, its parameters, search area or using an alternative object definition. Which alternative hypotheses are defined depends on the situation. The example in section 4.8.2 shows clearly that knowledge from the standards for road construction can be used for this purpose.

In the current implementation it is not possible that an object will get two classifications. Also the same object will not be segmented again, if it was already detected from another object. This situation will for example occur if a lane line is searched on a two lane road, because it is related to both edge lines. Fig. 4.27 shows the result of extracting the lane line from the aerial photograph of fig. 4.21. As can be seen only one line was segmented. This is done by checking first if the searched object type was already detected within the search area, before the segmentation starts.

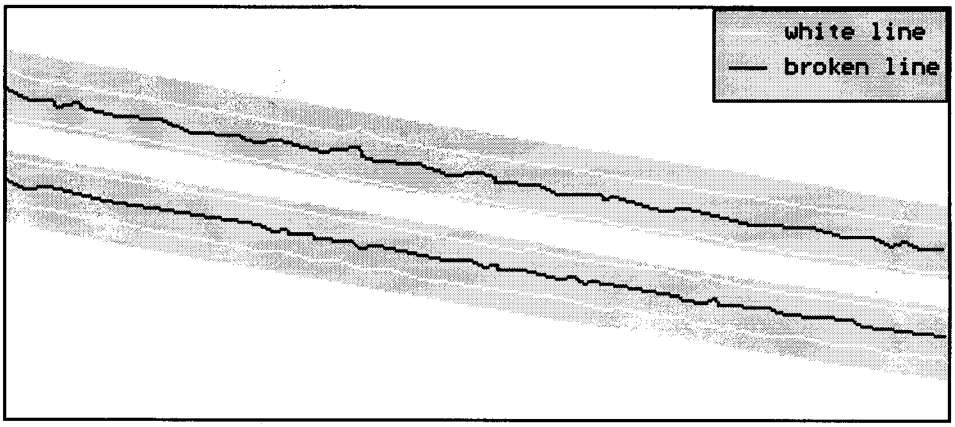

Fig. 4.27 Result of detection of one lane line related to two edge lines

Another common problem in reasoning is the danger of infinite recursion. In order to prevent this, identifications of searched objects are stored as related objects with the STM object.

In the next part of this thesis the presented concepts will be evaluated by concretizing them for a certain case and by performing experiments on multiple aerial images.

PART II

CASE STUDY: EXTRACTION OF NEW ROADS LINKED TO EXISTING MOTORWAYS

The aim of this part of the thesis is twofold. On the one hand to test if our research objectives are met: evaluate if knowledge-based concepts and the designed map-guided interpretation strategy are suitable for updating road maps. On the other hand to concretize the concepts for a certain case, which enables us to discuss details and implementation issues. This case study focuses on the recognition of new link roads connected to an existing motorway at medium scale. For this case experiments on multiple images were done to evaluate the research objectives. This imagery is used in the photogrammetric practice of the Survey Department of Rijkswaterstaat to produce large-scale road databases. Chapter 5 describes in detail the contents of the knowledge base, image processing techniques and control strategy. Next, in chapter 6, results are presented, discussed and evaluated.

CHAPTER 5

CONTENTS OF THE KNOWLEDGE-BASED AND THE DESIGNED INTERPRETATION STRATEGY

In this case study we focus our interest on the extraction of medium scale objects: road elements and junctions. The case is considered in which the outdated map only contains a motorway and that connected link roads present in the aerial photograph are newly constructed. Junctions on the main carriageway are changes which need to be detected, followed by extraction of the accompanying link roads using contextual reasoning. In this way the outdated map is updated corresponding to the new situation represented in the aerial photograph. The advantage of this case is that it can also be seen as an interactive application, in which the main road is indicated by a human operator.

This chapter describes details for this case, of which the results will be used in chapter 6 to evaluate the map-guided interpretation strategy designed in chapter 4. Specifically the next subjects are discussed:

- The contents of the knowledge base (section 5.3 and 5.4): which parts of the road network are defined as objects, which properties are used for recognition and how their values are determined, which standards for road construction are used and how are they incorporated;
- The interpretation strategy: which hypotheses are generated by the defined object relations (section 5.5), the image processing techniques activated for change detection and contextual reasoning and their parameter settings (section 5.6 and 5.7).

First, the objectives of the experiments are defined, because they determine the composition and characteristics of the image set (section 5.2) and contents of the knowledge base.

5.1 OBJECTIVES

Besides evaluation of the knowledge-based concepts and the designed interpretation strategy in general, our more specific objective in these experiments is to test whether incorporation of the specialized knowledge from the standards for road construction can improve the performance. Therefore, results are compared with those obtained when using general road properties. Although speed is not used directly as performance criterium, it is considered in the sense that experiments are done on images of reduced resolutions, because they are more suitable for development of the interpretation strategy, especially for tuning parameters of image processing routines and the composition of the contents of the knowledge base. As a result the influence of the resolution on the performance should be investigated.
When testing the performance of the map-guided interpretation strategy discrimination should be made between the ability to detect new junctions and road elements properly and the ability to classify them correctly as for example Y-junction or fly-over.

Summarizing, the objective of the experiments is to test the influence of the incorporation of standards for road construction against general road properties and the influence of resolution on detection and classification.

5.2 INPUT DATA

Aerial images and road databases are the input for the map-guided interpretation strategy. Its performance is tested in chapter 6 on multiple images of different situations. In this section the composition of the image data set and the contents of the road databases will be discussed.

5.2.1 IMAGE SET

The imagery used in this case study is from the photogrammetric practice of the Survey Department of Rijkswaterstaat. All aerial photographs depict motorways in the Western part of the Netherlands at scale 1:4000. Parts containing an intersection were selected. The fact if the Y-junction of a link road is present within the limited range of the photograph is used as criterium for selection. The photographs are located on parts of three different motorways:

- A12, between Utrecht and The Hague
- A4, between The Hague and Amsterdam
- A44, between the intersection with the A4 and Leiden

They are taken during three different photo flights.

***Fig. 5.1** Aerial image of T-intersection on 2-lane motorway A12, part of learning set*

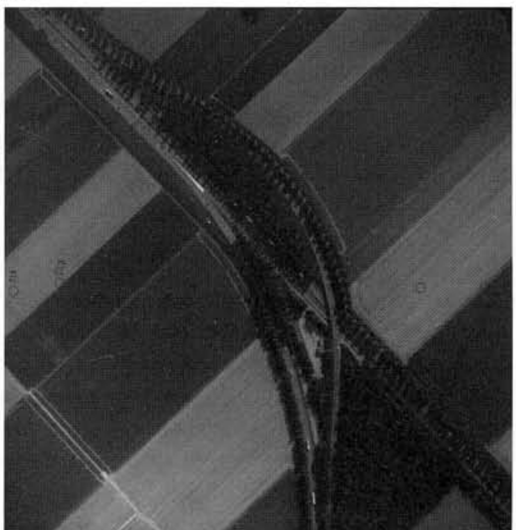

***Fig. 5.2** Aerial image of T-intersection on 2-lane motorway A4, part of learning set*

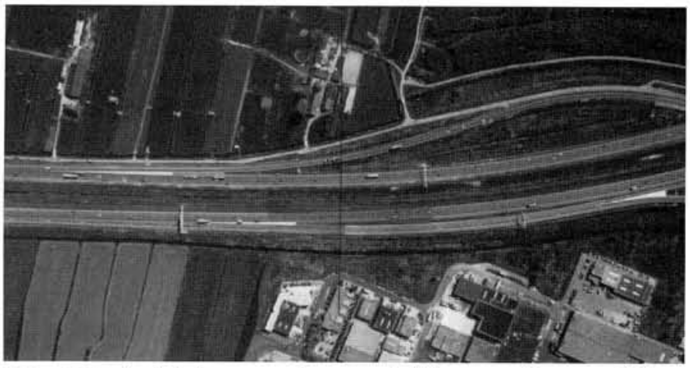

Fig. 5.3 *Aerial image of part of T-intersection on 3-lane motorway A4, part of learning set*

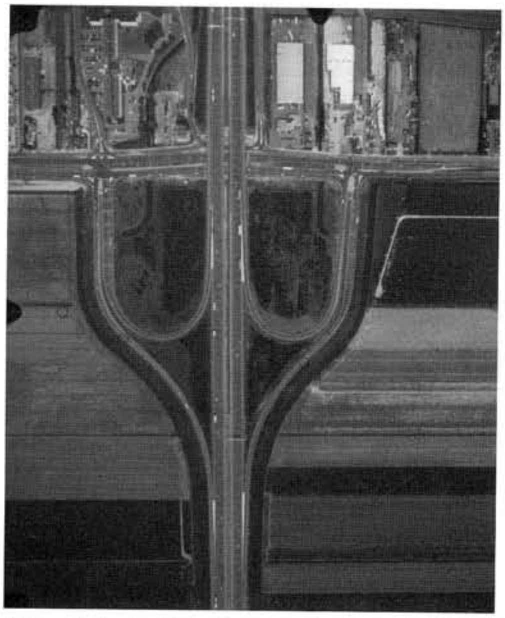

Fig. 5.4 *Aerial image of semi-cloverleaf interchange of the 4-lane motorway A4, part of learning set*

Fig. 5.5 *Aerial image of a interchange with Haarlemmermeer solution at the 2-lane motorway A12, part of learning set*

From the nine available scanned aerial photographs five were used to optimize the interpretation strategy, called the learning set. They are shown in fig. 5.1 - 5.5. Only these images were used to determine the values for conditions in the knowledge base and parameters for image processing. The remaining four images were used for testing and evaluation. The images of this test set are shown in fig. 5.6 - 5.9.

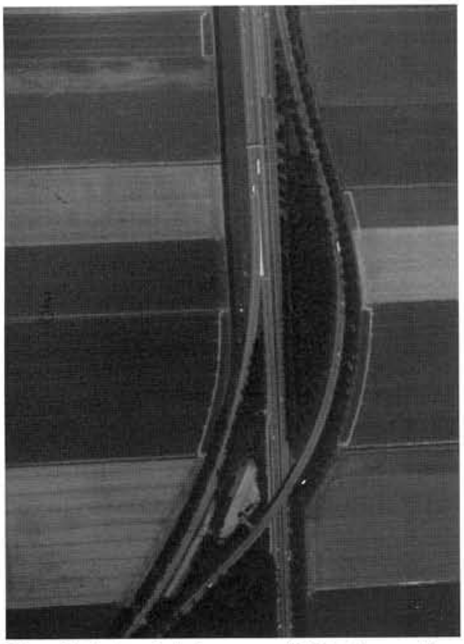

Fig. 5.6 Aerial image of T-intersection on 2-lane motorway A4, part of test set

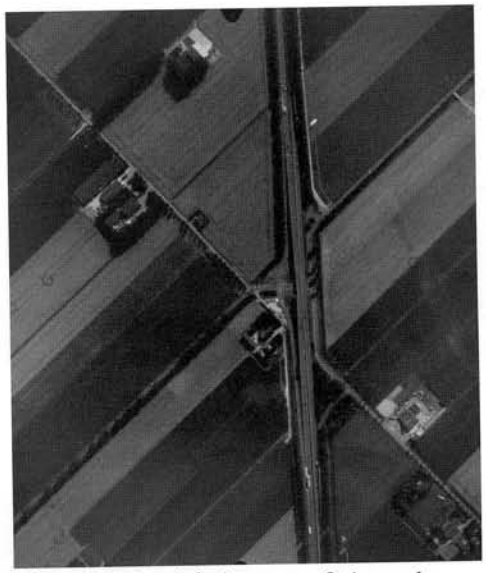

Fig. 5.7 Aerial image of interchange with Haarlemmermeer solution at the 2-lane motorway A44, part of test set

In all selected photographs the motorway consists of a dual carriageway. A condition for the learning set is that it is composed such that it contains images with 2-, 3- and 4-lane carriageways. Because there are far less motorways with 3 and 4 lanes than with 2 lanes, the learning set contains one image of 3-lane (fig. 5.3), one of 4-lane (fig. 5.4) and three of 2-lane carriageways (fig. 5.1, 5.2 and 5.5). These carriageways are part of two of the three different motorways: the A12 and A4. Another condition for the learning set is that it contains the all possible different types of intersections. Three main types (see section 4.3.1) are present:
- T-intersection (fig. 5.1, 5.2, 5.6)
- interchange, with Haarlemmermeer solution (fig. 5.5, 5.9) or semi-cloverleaf (fig. 5.4)
- part of T-intersection (fig. 5.3, 5.7) or interchange (fig. 5.8)

Fig. 5.2 and 5.6 depict the same T-intersection with and without shadow. Fig. 5.2, with shadow, was chosen to be part of the learning set, while fig. 5.6 is part of the test set. Fig. 5.6-5.8 are taken subsequently during the same photo flight and partly overlap each other.

Photographs were scanned with 100 μm pixel size and 8 bits per pixel, yielding a ground resolution of 0.40 m. per pixel. Images with a lower resolution were obtained using the simplest type of image pyramid construction [Rosenfeld, 1984]: averaging the grey values in non-overlapping 2-by-2 blocks of pixels. Repeating this method two times yields images at a resolution of 1.60 m. The motivation for using the lower resolution is that experiments with the original resolution are very time-consuming. Therefore, most experiments are carried out on low resolution and the best setting is compared with results at the original resolution. Because the photographs are in colour, they are scanned in red, green and blue. Intensity images are generated by averaging the values from these three bands. These images are used for the experiments.

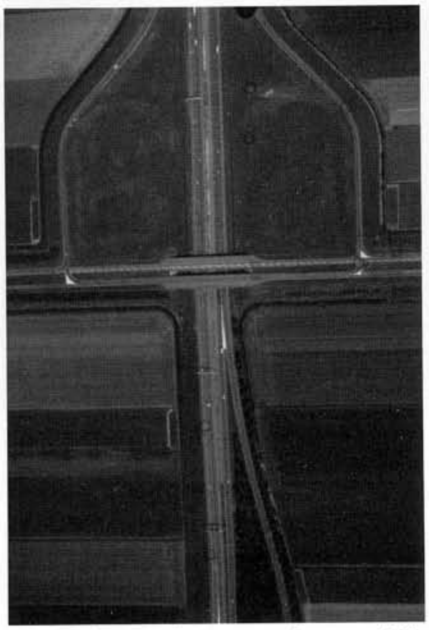

Fig. 5.8 Aerial image of part of T-intersection on 2-lane motorway A4, part of test set

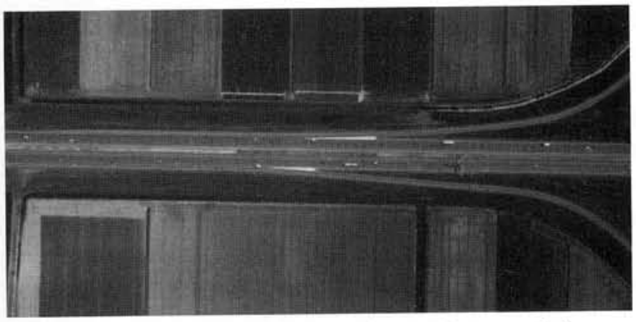

Fig. 5.9 Aerial image of part of interchange of the 4-lane motorway A44, part of test set

5.2.2 ROAD DATABASE

Because the present digital topographic database of the Survey Department represents objects not as a whole, but only as line elements, the needed road elements cannot be selected directly. Since their plan for the future is to convert to an object-oriented database, it was decided to extract the road elements manually from the aerial images.

The roads are extracted from the up-to-date aerial images as if they represent an outdated situation without link roads. If a slip-road is present, not the real road side is outlined, but a line as if there is a hard shoulder instead of this lane. As an example in fig. 5.10 the corresponding road database is projected as a raster image on the aerial photograph of fig. 5.1. For each aerial image the road database contains two road elements: the separated carriageways of a motorway.

Manual extraction results in coordinate lists, representing the contour of road elements. Before image interpretation can start, these need to be converted to STM objects. STM objects (see section 4.6.4) contain both an iconic description, i.e. the raster image of the road element, as well as a vector description, i.e. the measured coordinate list. The complete contour is split up into coordinate lists representing the side of the road and the border of the image. These are related to the original road element by part-whole relationships. The STM objects are not classified in advance, but obtain their classification during image interpretation. If they will be classified simply as "road element from database" or if discrimination will be made between 2-, 3- and 4-lane carriageways depends on the object definitions in the knowledge base, which will be subject of the next section.

Fig. 5.10 *Digital map of the motorway in the outdated situation projected on the aerial image in black*

5.3 CHOICE OF OBJECTS

In this section the objects are listed, which are defined in the knowledge base for this case study. Because one of our objectives is to test if incorporation of the standards improves the performance of the interpretation strategy, two different sets of objects are defined. One set in which the conditions for recognition are based on the standards for road construction [Rijkswaterstaat, 1975] and another set in which general road properties are used. Inclusion of the standards also leads to discrimination of more types of objects, because the standards contain detailed information about properties of specializations of road elements, like 2-, 3- and 4-lane carriageways of motorways. The terminology of section 4.3.2 (see also appendix B) is adopted, in order to discriminate at medium scale generalized from specialized object classes. The definitions of the objects in the specialized road network model are based on information from the standards, while the definitions of the objects in the generalized road network model are based on general road properties. This section describes which objects are chosen. Their properties, used to recognize these objects, are described in section 5.4. The results of interpretation using each of these road models are compared in chapter 6.

5.3.1 GENERALIZED ROAD NETWORK MODEL

In the generalized model no discrimination is made between the main carriageways and link roads. Both are classified as road elements and have the same general properties. The only difference is that one originates from the database and the other from segmentation. Junctions are defined at every position where two road elements are connected. They are seen as build from two parts: a part of a road element from the database which changed into a junction and the first part of a newly connected road element (fig. 5.11). In this case study all junctions originate from the segmentation, since they are not present in the databases and need to be detected.

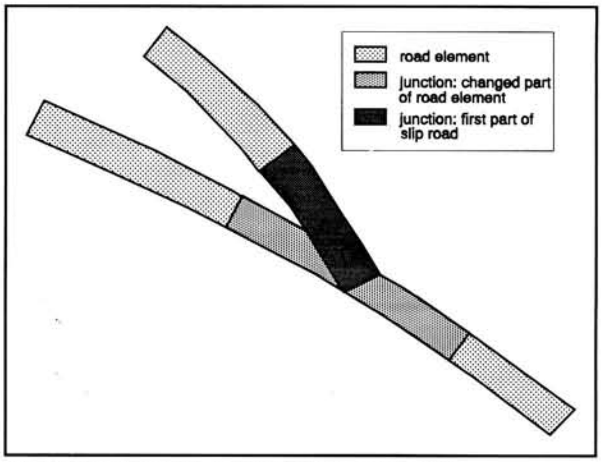

Fig. 5.11 Y-junction build from two parts

In conclusion the next medium scale objects are defined in the generalized model:
- road element from database
- junction from segmentation
- road element from segmentation

5.3.2 SPECIALIZED ROAD NETWORK MODEL

In the specialized model all road elements from the database are dual main carriageways, part of the specialized type "motorway" at small scale. They are divided into main carriageways with 2, 3, or 4 lanes. Single main carriageway could be added if needed. The specialized type of the road in the database is not directly specified and needs to be inferred during the interpretation process.
The new road elements, which need to be extracted by segmentation, are link roads. Discrimination is made between link roads with 1 lane and 2 lanes. Link roads branch off specializations of junctions, in particular Y-junctions. These are again seen as build from two parts: a changed part of a main carriageway and a part of a link road. The advantage of this viewpoint is that it is easier to evaluate the geometric properties of Y-junctions, because both parts have their own properties. The part of the main carriageway has a width related to the fact whether it consists of 2, 3 or 4 lanes, while the width of the part of the link road corresponds to the incorporation of 1 lane or 2 lanes. Also the angle between both parts, which can be used to discriminate y-junctions and fly-overs, can more easily be measured. Besides, this part of a link road can be utilized to the further track the link road during contextual reasoning (see section 5.6.2).

The other specialization of junctions which are discriminated are fly-overs. Also fly-overs are seen as build from two parts.

Summarizing, the specialized road network model consists of the next seven objects:
- main carriageway with 2, 3 or 4 lanes
- link road with 1 or 2 lanes
- Y-junction
- fly-over

5.4 DEFINED CONDITIONS FOR RECOGNITION OF OBJECTS

In the knowledge base road features are defined as conditions for the recognition of segmented objects. Geometric, radiometric and spatial conditions used for recognition of the objects in section 5.3 are discriminated. These conditions are subsequently described in this section and for each condition it is indicated for the recognition of which object of the generalized and/or specialized model it is used. For each condition a range of allowed values, attached to it, is determined. The images in the learning set are used to derive experimentally suitable values. For objects from the specialized road network model also the standards for road construction [Rijkswaterstaat, 1975] are used for this purpose.

5.4.1 GEOMETRIC CONDITIONS

Two geometric conditions are used in the knowledge base: average width and the L/W-ratio.

Average width

In the specialized road model the average width is added as geometric property to discriminate the various specialized types. The values are derived from the widths of traffic areas on the carriageway as defined in the standards for road construction [Rijkswaterstaat, 1975]:

traffic lane:	3.5 m.
hard shoulder:	3.0 m.
correction strip:	1.1 m. (main carriageway)
	0.6 m. (link road)
line mark:	0.15 m.

They are compared with the values for the average width measured in the images of the learning set in table 5.1. The width is measured at many positions along every road. The number of measurements for calculation of the average width and the number of main carriageways present in the learning set are indicated as well in table 5.1.

type of road element	standard width	measured average width	standard deviation of average	number of carriageways	number of measurements
2-lane main carriageway	11.55 m.	11.95 m. (11.50 m.)	1.33 m. (1.36 m.)	6 (4)	222 (138)
3-lane main carriageway	15.20 m.	15.43 m.	0.47 m.	2	128
4-lane main carriageway	20.75 m.	19.75 m.	0.45 m.	2	73
1-lane link road	7.40 m.	7.43 m.	1.46 m.	5	220
2-lane link road	11.05 m.	11.01 m.	1.27 m.	5	298

Table 5.1 *Measured average width and width from standards for road construction [Rijkswaterstaat, 1975]. The values in brackets are obtained when a 2-lane carriageway which becomes a 4-lane carriageway is ignored in the calculations.*

There is only one photograph which contains 3- and 4-lane main carriageways. As a result the calculation of this average width deviates considerably from the standard width. The standard deviation is very small and has a magnitude of about one pixel (0.40 m.). This is obviously the measuring accuracy, but does not say anything about the deviation of the width of carriageways in relation to the standard. The standard deviation of the measured average width based on more carriageways is larger: between 1.25 and 1.50 m. But, for the link roads the measured average width only slightly differs from the standard width. A remark is that the link roads in one photograph were not included in the calculation of the average width, because they started as 1-lane link roads and ended as 2-lane link roads. Maybe a separate class should be defined for this type of link roads. The average width of the 2-lane main carriageway differs more from the standard than for link roads, but it turns out that in one of the photographs the 2-lane carriageway partly has two hard shoulders, because a bit further, outside the range of the photographs, it becomes a 4-lane carriageway. If this 2-lane motorway is ignored during calculation of the average, the average measured width becomes 11.50 m., which is very close to the standard width.

In conclusion, the width according to the standards for road construction is suitable to define the interval (table 5.2) within which the average width of a segmented object should lie to be a certain specialized type of road element. Assuming a normal distribution a definition of a 95% confidence interval yields an interval of about 2.5 - 3.0 m. at both sides of the standard width, according to the measured standard deviation. However, in order to prevent ambiguity in interpretation, the intervals are not allowed to overlap. Consequently, we are forced to use the largest possible interval with the boundary in the middle between the standard widths (W_{std}) of 2- and 3-lane main carriageways and 1- and 2-lane link roads, yielding an interval [W_{std}-1.825 m., W_{std}+1.825 m.] The same interval was used for 4-lane main carriageways for reasons of consistency, resulting in a small gap between the intervals for 3- and 4-lane roads. For the minimum width of a 1-lane link road and the maximum width of a 2-lane link road it was decided not to hold on to a symmetrical interval, because there are several link roads with different features within one class. For example: some link roads do not have a hard shoulder at one side, as also reported in the standards for road construction [Rijkswaterstaat, 1975], but have a correction strip at both sides. Also according to the standards the width of link roads is enlarged to maximal 4.5 m. in curves with a radius smaller than 300 m. These two cases are used for determination of upper and lower limit of the interval of the link roads.

type of road element	minimum width	maximum width
2-lane main carriageway	9.7 m.	13.7 m.
3-lane main carriageway	13.7 m.	17.0 m.
4-lane main carriageway	18.9 m.	22.6 m.
1-lane link road	4.7 m.	8.4 m.
2-lane link road	8.4 m.	14.5 m.

***Table 5.2** Range of width for recognition of specialized types of road elements*

L/W-ratio

The L/W-ratio is defined as the average length of the segmented object divided by its average width. It is utilized to select only those parts of road elements from the database which possibly changed into a junction and to reject those which are detected due to e.g noise. In the generalized road network model a condition for the length of these parts, measured along the road, is that it should be larger than the width of the road, because it corresponds at least to the width of the crossing road, in case of a perpendicular crossing. Thus the L/W-ratio is larger than 1. The width of the crossing road is assumed to be at least as wide as one carriageway of the main road. In the specialized model the minimum length of changed parts is equal to the width of a crossing 1-lane link road, yielding L/W-ratios of 0.64, 0.49 and 0.36 for 2-, 3-, and 4-lane carriageways respectively, using the standard widths of table 5.1.

In the specialized road network model an additional task of the L/W-ratio is to discriminate fly-overs from Y-junctions. Table 5.3 shows the L/W-ratios for detected parts of road elements at the location of a fly-over or Y-junction in the learning set at a resolution of 1.60 and 0.40 m. L/W-ratios of the same junction at different resolutions are given on one line. As can be seen L/W ratios at high resolution are in general larger. Consequently at each resolution a different threshold for the L/W ration was chosen to discriminate fly-overs and Y-junctions. At neither resolution it is possible to choose such a value for the threshold between both classes that always a correct classification is obtained. For this reason an additional spatial condition is defined in section 5.4.3. A value of 6.30 was chosen at a resolution of 1.60 m., because it yields only one misclassified Y-junction. It belongs to a location of a Y-junction at which two small changed parts were detected instead of one. One of them is included in table 5.3 and is marked with an asterisk. A value of 7.00 is chosen at a resolution of 0.40 m., in which case the two marked fly-overs of the learning set get an incorrect classification.

Fly-over		Y-junction	
1.60 m.	0.40 m.	1.60 m.	0.40 m.
3.78	8.02 *	50.32	52.52
2.35	2.96	22.38	21.31
6.09	8.42 *	11.26	12.61
6.13	6.55	8.45	9.06
5.53	6.26	6.49	8.22
		19.15	20.25
		3.58 *	9.18
		9.44	7.64
		11.69	11.68
		29.50	28.37

Table 5.3 *L/W-ratio for fly-overs and Y-junctions at a resolution of 1.60 and 0.40 m. Values marked with an asterisk will be classified incorrectly with the chosen threshold for the L/W ratio of 6.30 and 7.00 respectively.*

5.4.2 RADIOMETRIC CONDITIONS

A property of road elements is their characteristic profile of grey values perpendicular to the road. An example of a set of profiles is given in fig. 5.12. An extracted profile contains road pixels and some background pixels. Usually road pixels in intensity images have a higher grey value than background pixels. By comparing an extracted grey value profile with a pre-defined profile model, road elements can be recognized. A profile model, which includes grey values from the aerial image, is used for searching road elements. An artificial profile model is used to obtain the grey-value based profile model.

In the artificial profile model it is assumed that the grey value of the road is higher than its surrounding. A raster image of the road element and its near road elements present in the data-base is drawn with value 255 for the road and 0 for the background. This raster image and the aerial image are both resampled perpendicular to the axis of the considered road element. Each artificial profile is matched with the corresponding grey level profile.
The grey-value based profile model is generated by calculating the mean grey-level profile using

only those profiles which match best with the artificial profile model. The percentage of best matches which is used to calculate the average is a parameter of the image processing technique for change detection. An advantage of this model is that real grey values are included and consequently variations due to differences in exposure of the photographs and overall features, like linear road markings, are included in the model. However, the local presence of other road elements is averaged. So other road elements, like a parallel carriageway, are only included if they are constantly present along the complete length of the road element.

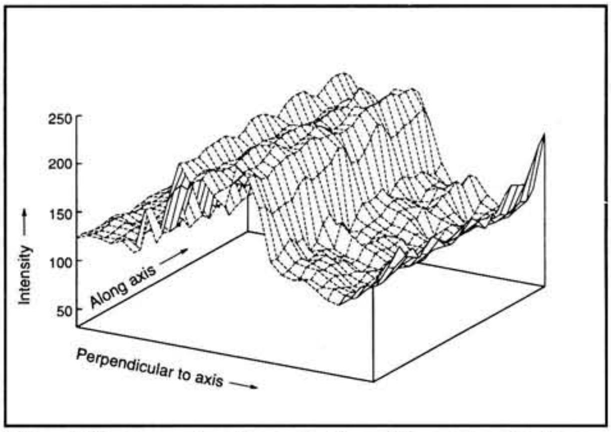

Fig. 5.12 *Example of a set of profiles extracted from an aerial image*

The cross-correlation is used as criterium to judge the correspondence between an extracted grey value profile and the profile model. This criterium is passed as a parameter attached in the knowledge base to the image processing technique for changed detection. The value used as threshold for the cross-correlation will be derived in section 5.7.

5.4.3 Spatial conditions

One of the spatial conditions which are used in the knowledge base to distinguish Y-junctions from fly-overs is the angle between the two parts of a junction. A Y-junction usually indicates a sharp angle, while at a fly-over roads can even cross perpendicular. According to the standards for road construction [Rijkswaterstaat, 1975], the angle in a Y-junction is between 0.6 and 5.1 degrees, preferable 1.7 degrees. However, the angles between the detected parts of the Y-junctions measured in the learning images are larger. The reason is that often not the first part of the link road is found, but a part further from the branchpoint, where the link road already bends off more. Therefore, the measured values (table 5.4) were used to determine the condition for the angle of both junctions instead of the values from the standards for road construction. Another remarkable fact is that the angles of fly-overs in table 5.4 are rather small. Perpendicular crossing are expected at a fly-over. However, these are not found because in all cases they end into a crossing very close to the main carriageway. If the link road is not found, the angle cannot be measured.
Again it is not possible to determine an unambiguous value for the threshold between both classes. But overlapping intervals can be defined, because this is the second condition for discrimination between fly-overs and Y-junctions. For Y-junctions an interval between 0 and 21 degrees is defined. All angle between link roads and road elements detected in the learning set will be accepted. For fly-overs a range between 18 and 90 degrees is defined. One angle is

rejected, but it corresponds to the angle between a link road and a part of a road element, which was already misclassified, because of its L/W-ratio. The same intervals are used for both resolutions.

Angle fly-over (in degrees)		Angle Y-junction (in degrees)	
1.60 m.	0.40 m.	1.60 m.	0.40 m.
29.78	32.97	3.28	-
19.21	20.00	10.27	9.93
28.94	17.89 *	10.59	10.49
		9.40	-
		-	10.03
		18.81	16.18
		17.25	15.78
		20.01	14.45
		6.48	6.72

Table 5.4 *Angles in degrees between main carriageway and link road for fly-overs and Y-junctions*

5.5 INTERPRETATION STRATEGY

Updating the situation in this case study requires change detection as well as contextual reasoning. The interpretation strategy is established by the defined temporal and spatial relationships between the objects. The object relationships for objects of the generalized and specialized road network model respectively are listed in the next sections.

5.5.1 RELATIONS BETWEEN GENERALIZED OBJECTS

Representation of the developed interpretation strategy as a semantic network results for the objects of the generalized road network model in fig. 5.13. Hypotheses generated for the interpretation strategy correspond to object relations in the semantic network. The frames corresponding to these three relationships are represented in fig. 5.14 - 5.16.

```
DEFINE RELATION
    type:   changed into
    from:   road element
    into:   part of junction

// part for hypothesis generation
    priority:    1
    optional: YES

// part for goal-directed segmentation:
    image processing technique:
    - profile matching
    parameters:
    - % for model option
    - threshold for cross-correlation
    - factor for width search area
    search area: centred around the road
    element
ENDDEF
```

Fig. 5.14 Definition temporal relationship

```
DEFINE RELATION
    type:   spatially related to
    from:   part of junction
    to:     junction

// part for hypothesis generation
    priority:    2
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - rotating profile analyser
    parameters:
    - position on axis
    - range of distance to search area
    - length of search area
    - width of search area
    - range of width slip road
    - range of rotation
    - angle difference
    - threshold for cross-correlation
    search area: at the other side of the
    junction than the parallel carriageway
    runs
ENDDEF
```

Fig. 5.15 Definition of spatial relationship

```
DEFINE RELATION
    type:   spatially related to
    from:   junction
    to:     new road element

// part for hypothesis generation
    priority:    3
    optional: YES

// part for goal-directed segmentation:
    image processing techniques:
    - extrapolating profile analyser
    parameters:
    - step size
    - width profile model
    - deviation from predicted position
    - threshold for cross-correlation
    search area: on top of junction
ENDDEF
```

Fig. 5.16 Definition of spatial relationship

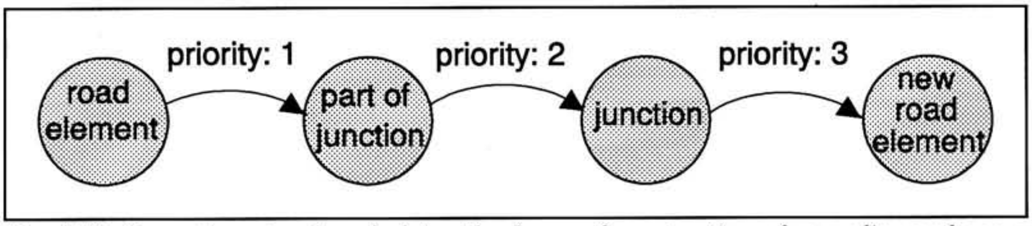

Fig. 5.13 Semantic network underlying the frames for extraction of new slip roads on a motorway using objects of the generalized road network model

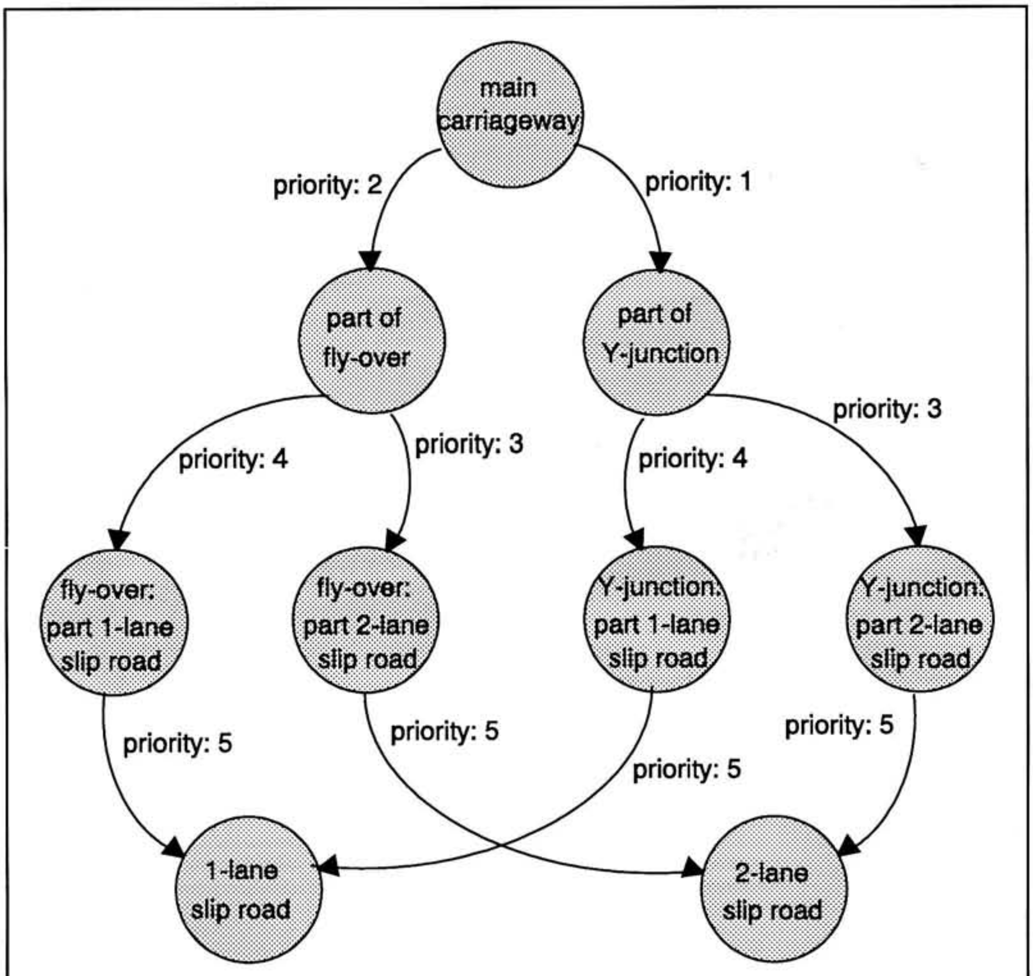

Fig. 5.17 Semantic network underlying the frames for extraction of new slip roads on a motorway using the objects of the specialized road network model

The increasing priority of the relationships between road elements from the database, junctions and new road element means that hypotheses for detection are successively generated. This results in a rather straightforward process of change detection followed by contextual reasoning. The first object relationship (fig. 5.14) generates hypotheses for change detection: road elements from the database are searched that possibly changed into a junction. The other two frames define spatial relationships for contextual reasoning. The first (fig. 5.15) aims to detect the first part of a new road element, which forms together with the changed part a new junction. This is an essential relationships, which means that if it is not found, the changed part is rejected as part of a junction. The next relationship (fig. 5.16) activates the image processing technique to track the new road element till it ends, usually at the image border. The image processing techniques activated by these relations and their parameters and search area will be discussed in detail in section 5.6 and 5.7.

5.5.2 Relations between specialized objects

The semantic network underlying the relations between objects of the specialized road network model (fig. 5.17) is more extensive and more complex. Actually there are even more objects and relationships present in the knowledge base, because discrimination is made between main carriageways with 2, 3 and 4 lanes. For each class a semantic network similar to the one presented in fig. 5.17 is defined. The frames corresponding to the object relationships in the semantic network for main carriageways in general are represented in fig. 5.18 - 5.27.

```
DEFINE RELATION
    type:   changed into
    from:   main carriageway
    into:   part of Y-junction

// part for goal generation
    priority:   1
    optional: YES

// part for goal-directed segmentation:
    image processing technique:
    - profile matching
ENDDEF
```

Fig. 5.18 Definition temporal relationship

```
DEFINE RELATION
    type:   changed into
    from:   main carriageway
    into:   part of fly-over

// part for goal generation
    priority:   2
    optional: YES

// part for goal-directed segmentation:
    image processing technique:
    - profile matching
ENDDEF
```

Fig. 5.19 Definition temporal relationship

In the interpretation strategy generated by these object relationships first, parts of the carriageway are searched which changed into a Y-junction (fig. 5.18) and next, those which changed into a fly-over (fig. 5.19). After change detection in each of these hypothesized parts of a junction first, the beginning of a 2-lane link road is searched (fig. 5.20 and 5.21) and next, of a 1-lane link road (fig. 5.22 and 5.23) in order to form a complete Y-junction or fly-over. Because these relationships are essential, the changed parts are rejected if the first part of a link road cannot be found. Finally, the complete link roads are searched by tracking starting at the Y-junction or fly-over (fig. 5.24-5.27). The parameters and search area of the image processing techniques are the same as in fig. 5.14-5.16, but are omitted for convenience of arrangement. Compare fig. 5.18 - 5.19 with fig. 5.14, fig. 5.20-5.23 with 5.15 and fig. 5.24 - 5.27 with 5.16.

```
DEFINE RELATION
    type:   spatially related to
    from:   part of Y-junction
    to:     Y-junction with part of 2-lane
            link road

// part for hypotheses generation
    priority:   3
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - rotating profile analyser
ENDDEF
```

***Fig. 5.20** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   part of fly-over
    to:     fly-over with part of 2-lane link
            road

// part for hypotheses generation
    priority:   3
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - rotating profile analyser
ENDDEF
```

***Fig. 5.21** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   part of Y-junction
    to:     Y-junction with part of 1-lane
            link road

// part for hypotheses generation
    priority:   4
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - rotating profile analyser
ENDDEF
```

***Fig. 5.22** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   part of fly-over
    to:     fly-over with part of 1-lane link
            road

// part for hypotheses generation
    priority:   4
    essential: YES

// part for goal-directed segmentation:
    image processing techniques:
    - rotating profile analyser
ENDDEF
```

***Fig. 5.23** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   Y-junction
    to:     1-lane link road

// part for hypotheses generation
    priority:   5
    optional: YES

// part for goal-directed segmentation:
    image processing techniques:
    - extrapolating profile analyser
ENDDEF
```

***Fig. 5.24** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   Y-junction
    to:     2-lane link road

// part for hypotheses generation
    priority:   5
    optional: YES

// part for goal-directed segmentation:
    image processing techniques:
    - extrapolating profile analyser
ENDDEF
```

***Fig. 5.25** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   fly-over
    to:     1-lane link road

// part for hypotheses generation
    priority:   5
    optional: YES

// part for goal-directed segmentation:
    image processing techniques:
    - extrapolating profile analyser
ENDDEF
```

***Fig. 5.26** Definition of spatial relationship*

```
DEFINE RELATION
    type:   spatially related to
    from:   fly-over
    to:     2-lane link road

// part for hypotheses generation
    priority:   5
    optional: YES

// part for goal-directed segmentation:
    image processing techniques:
    - extrapolating profile analyser
ENDDEF
```

***Fig. 5.27** Definition of spatial relationship*

5.6 IMAGE PROCESSING TECHNIQUES AND THEIR PARAMETERS

Attached to each object relation are image processing techniques which will be activated to segment the searched object type. In this section the image processing techniques for change detection and contextual reasoning together with their parameters will be described in detail. Setting these parameters is the subject of section 5.7.

5.6.1 CHANGE DETECTION

The image processing technique described in this section aims at detecting whether there are parts of the medium scale road element from the map which changed into a part of a junction. It applies so-called signal or area-based matching [Lemmens, 1988] of a profile model and one-dimensional grey-level profiles extracted from the aerial photograph at the position of the road element in the map. In order to obtain grey-level profiles, the aerial image is resampled perpendicular to the local direction of the axis of the road element in the map. Every extracted grey-level profile is matched with the grey-level based profile model, described in section 5.4.2. The percentage of best matches (B%) used for the calculation of the average grey-level profile is a parameter. Cross-correlation is used as criterium for matching and its threshold (R_C) is passed as parameter as well. Parts of road elements which correspond with groups of profiles with low values for the cross-correlation are returned to be evaluated as road elements which changed into a junctions.

The search area is defined centred around the road element from the database with a width related to the mean width of the road element (W_R). This width also determines the width of the profile model (W_S). The factor that establishes this relation is passed over as one of the parameters of the image processing routine.

This image processing technique is attached to the relation in fig. 5.14 between generalized objects and the relations in fig. 5.18 and 5.19 between specialized objects.

5.6.2 Contextual reasoning

Contextual reasoning is used for two purposes:
- detection of the part branching off a changed part of a road element to form a junction;
- tracking of link roads starting at a detected new junction.

Detection of junctions

The image processing routine for the detection of junctions is based on modelling the junction as a straight line starting from the axis of the main carriageway at the position of a changed road element. At the top of this line a rectangular search area is defined in which a series of characteristic profiles needs to be found in order to recognize the line as part of a junction (fig. 5.28). As should be noted the real shape of the junction is not detected.

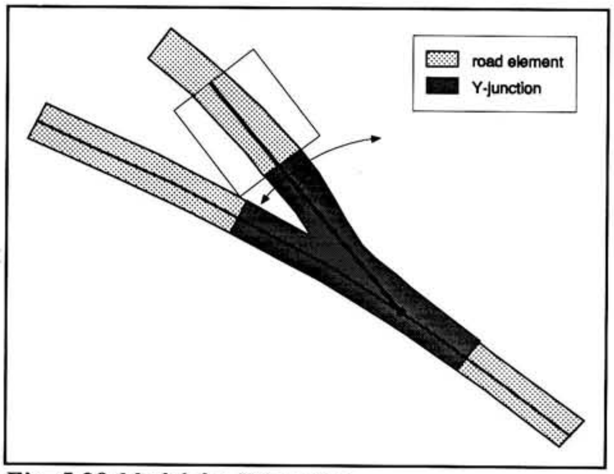

***Fig. 5.28** Model for Y-junction*

Within the search area the aerial image is resampled perpendicular to the predicted line. The resampled image can be seen as a series of one-dimensional grey-value profiles. Next, area-based matching of this resampled image with a series of artificial profiles is performed. The criterion for a good match is again defined by the cross-correlation. Since the width of the link road is not exactly known, several values for the width of the road in the artificial road model are tested. Also the direction of the line is not exactly known and therefore gradually adapted to search the Y-junction. The optimal value for the distance to the search area varies for different types of intersections like semi-cloverleafs and Haarlemmermeer solutions. All distances within a predefined range are tested. Standards for road construction are used to restrict the range of values for width, distance and direction which are tested. Since only one new road element is assumed to start at each changed road element, the best match from all varied directions, distances and widths is returned to be evaluated as Y-junction.

The image processing procedure for detection of junctions has many parameters of which some are related to each other. Figure 5.29 visualizes the unknown variables of the algorithm. Two groups can be distinguished:

1) the geometric variables determine at which positions the beginning of a link road is searched:
   a) position on the main road (x,y)
   b) distance to the search area D_i
   c) angle with the main road φ_i
2) the variables which determine the size of search area and model for local evaluation of the presence of a junction by radiometry:
   a) length of the search area L_S (in direction of line)
   b) width of the search area W_S (perpendicular to line)
   c) width of the road in the model M_i

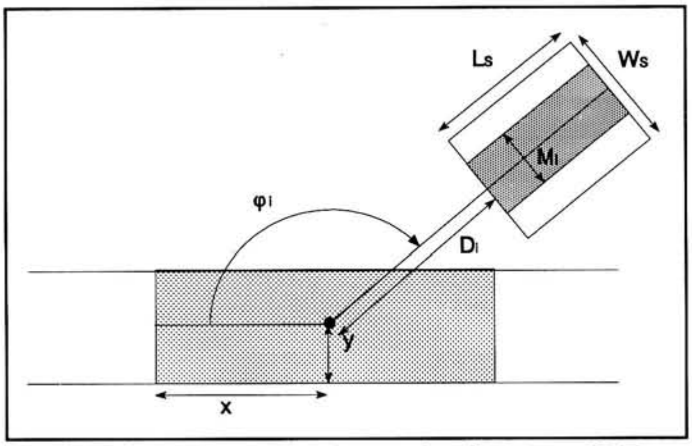

***Fig. 5.29** Parameters of technique for detection of Y-junctions*

Both groups have influence on the value of the criterion for the presence of a link road, the cross-correlation R_C, which is also a parameter.

The variables 1b, 1c and 2c are varied, resulting in the next parameters:

$[D_{min}, D_{max}]$	range within which the distance from the starting point on the axis to the road element to the beginning of the search area is varied
$[\varphi_{min}, \varphi_{max}]$	range within which the direction is rotated
$\Delta\varphi$	angle difference between successive directions which are tested

[M_{min}, M_{max}] range within which the width of the road in the model is varied
Fig. 5.30 and 5.31 visualize the variation of these parameters. Settings will be discussed in more detail in section 5.7.

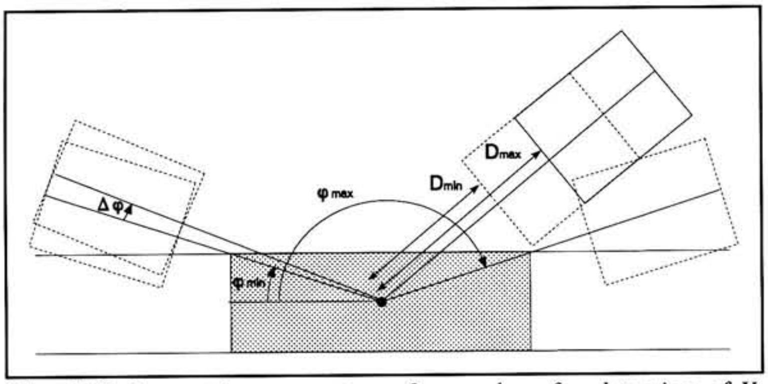

Fig. 5.30 *Geometric parameters of procedure for detection of Y-junction, which vary within a range*

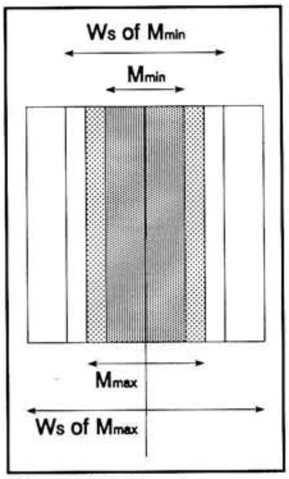

Fig. 5.31 *Varying parameter for width road in model*

Road tracking

The procedure for road tracking is based on profile analysis like in [Groch, 1982] and [McKeown/Denlinger, 1988]. The basic idea is evaluation of the one-dimensional grey level profile perpendicular to the road at a road point predicted by extrapolation of previously detected road points. In contrast to the procedure developed in [Gunst, 1991] every profile between the previously detected and predicted road point is evaluated instead of only one profile. In this way the algorithm is less sensitive for the presence of small disturbances on the road surface, e.g. due to cars, at the cost of more computation time.

Road tracking (fig. 5.32) starts with predicting a road point and resampling perpendicular to the predicted direction of the road. The similarity of an extracted profile (C) with a profile model (A-B) is a measure for the significance of the predicted point. An iterative convolution procedure searches the most significant point close to the predicted point. It calculates the root mean square difference between the model and the extracted profile. Therefore, the grey-level profile extracted from the aerial image is wider than the profile model such that the profile model can shift over the extracted profile in order to detect a deviation between the predicted and real position. The maximum deviation (Δp) is one of the parameters of the algorithm. By cross-correlation the similarity of profile (D) and the profile model is quantified. In fig. 5.32 intermediate extracted profiles are omitted for the purpose of giving a clearer illustration of the process, but the previous procedure is performed for every extracted profile.

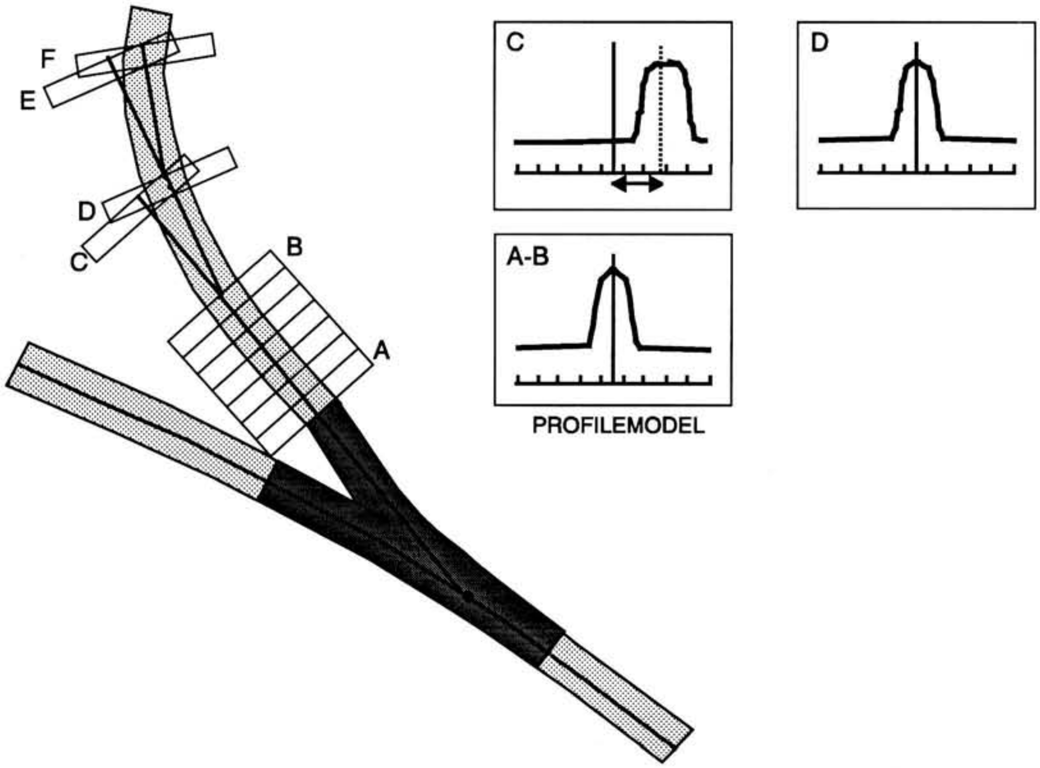

Fig. 5.32 Road tracking by profile analysis

The first part of the link road, detected as part of the junction, is used to obtain the initial profile model. An artificial profile model is used to detect local deviations in the positions of the straight road element used for detection of the junction. This is necessary because there is often a curve in the first part of a link road. By averaging all grey level profiles (A-B) on the road element, the initial profile model is obtained. The next road point is predicted by linear extrapolation from the last road point of the road element using the final calculated perpendicular direction after correction of this point. The distance to the predicted point is called the step size (s) and is related to the width of the road element detected in the junction. The idea is that narrow roads have a higher curvature, so the step size should be smaller. The step size is reduced till the last accepted profile if the farthest profiles show insufficient correspondence with the model. Intermediate rejected profiles can be evaluated as junction in a recursive process. If no profile within the search area is accepted, the step size is first doubled, assuming a large junction and if this still does not yield a road segment, tracking stops.

5.7 PARAMETER SETTINGS

In this section the used settings for the parameters of the image processing techniques described in section 5.6 will be given. Finding an optimal of all these parameters is very difficult, because most of them depend on each other. In appendix C.2 it is investigated how sensitive the parameter setting of the image processing procedure for detection of junctions is for variations in the result of the procedure for detection of changes.

5.7.1 SETTINGS OF ALL PARAMETERS USED IN THE CASE STUDY

Four approaches, sometimes combined, are used to determine the optimal settings:
1) mathematical or geometrical relations
2) experiments on artificial images
3) experiments on aerial images
4) measurements in database

Examples of how these methods are used for determination of values of parameters are given in section 5.7.2. Table 5.5 - 5.7 show the setting of all parameters and which of these methods is used to determine their values. Notations are listed in appendix D.

Some parameter settings differ for the generalized and specialized road network model. In the specialized model it is possible to use the same image processing technique with slightly different parameters for each object. A different criterium for the cross-correlation can for example be used for 2-, 3- and 4-lane motorways, because it depends on the width of the road. Standards for road construction [Rijkswaterstaat, 1975] are if possible included to determine parameter settings for the specialized case.

Parameters	GENERALIZED ROAD NETWORK MODEL		SPECIALIZED ROAD NETWORK MODEL		
	Method	Value	Method	Knowledge from standards	Value
B%	3	10%	3	-	10%
W_S	1	$2.0 \times W_R$	1	lane width	$2.0 \times W_R$
R_C	3	0.70	1+3	width 2-, 3- and 4-lane motorways	0.70 (2 lanes) 0.75 (3 lanes) 0.79 (4 lanes)

Table 5.5 Values of all parameters for change detection, the method used to determine them and the knowledge from the standards for road construction included in the specialized road network model

Parameters	GENERALIZED ROAD NETWORK MODEL		SPECIALIZED ROAD NETWORK MODEL		
	Method	Value	Method	Knowledge from standards	Value
(x,y)	1	$(\frac{1}{2}L_R, \frac{1}{2}W_R)$	1	-	$(\frac{1}{2}L_R, \frac{1}{2}W_R)$
φ_{min}	1	$\arctan(\frac{W_R}{L_R})$	1	-	$\arctan(\frac{W_R}{L_R})$
φ_{max}	1	$180° - \varphi_{min}$	1	angle in a Y-junction	Y-junction: $30° + \varphi_{min}$ at both sides (see fig. 5.33) fly-over: $180° - \varphi_{min}$ (see fig. 5.30)
$\Delta\varphi$	2	1.0	2	-	Y-junction: 0.5 fly-over: 1.0
D_{min}	1	$\frac{W_R + M_{min}\cos(\varphi_i)}{2\sin(\varphi_i)}$	1	width 1- and 2-lane link road	$\frac{W_R + M_{min}\cos(\varphi_i)}{2\sin(\varphi_i)}$
D_{max}	3	$1.0 \times L_R$	3	-	$1.0 \times L_R$
L_S	3	$3.0 \times \boldsymbol{w_r}$	1	maximum curvature link road (R_{min} = 35 m.) and width 1-/2- lane link road	$\sqrt{4M_iR_{min} - 3M_i^2}$ (see fig. 5.34)
W_S	1	$2.0 \times W_R$	1	width lane	$2.0 \times M_i$
M_{min}	3	w_r - 16 m.	4	width 1- and 2-lane link road	4.725 (1 lane) 8.375 (2 lanes) (see table 5.2)
M_{max}	3	w_r + 16 m.	4	width 1- and 2-lane link road	8.375 (1 lane) 14.50 (2 lanes) (see table 5.2)
R_C	3	0.75	3	-	0.75

Table 5.6 *Values of all parameters for detection of junctions, the method used to determine them and the knowledge from the standards for road construction included in the specialized road network model*

Parameters	GENERALIZED ROAD NETWORK MODEL		SPECIALIZED ROAD NETWORK MODEL		
	Method	Value	Method	Knowledge from standards	Value
s	3	$3.0\times M_W$	1	maximum curvature link road (R_{min})	$\sqrt{M_W R_{min} - \frac{3}{4} M_W^2}$ (see fig. 5.34)
W_S	1	$2.0\times M_W$	1	width lane	$2.0\times M_W$
Δp	3	10 m.	1	maximum curvature link road (R_{min})	$R_{min} - \sqrt{R_{min}^2 - s^2}$ (see fig. 5.34)
R_C	3	0.70	3	-	0.70

Table 5.7 *Values of all parameters for road tracking, the method used to determine them and the knowledge from the standards for road construction included in the specialized road network model*

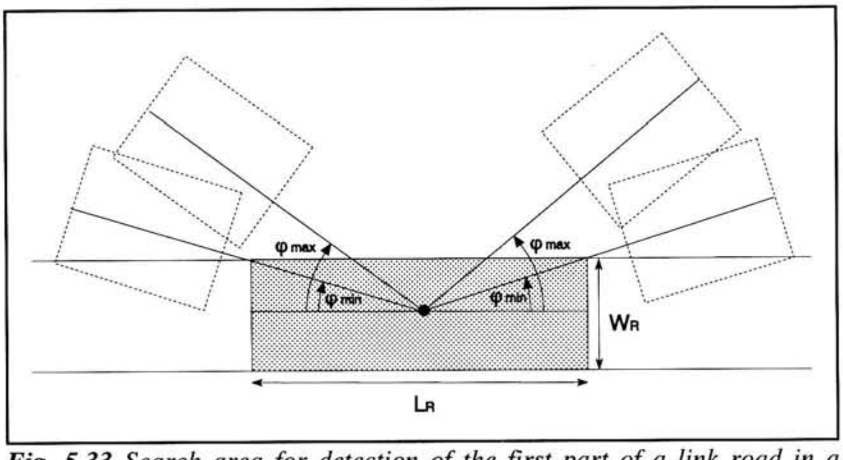

Fig. 5.33 *Search area for detection of the first part of a link road in a Y-junction*

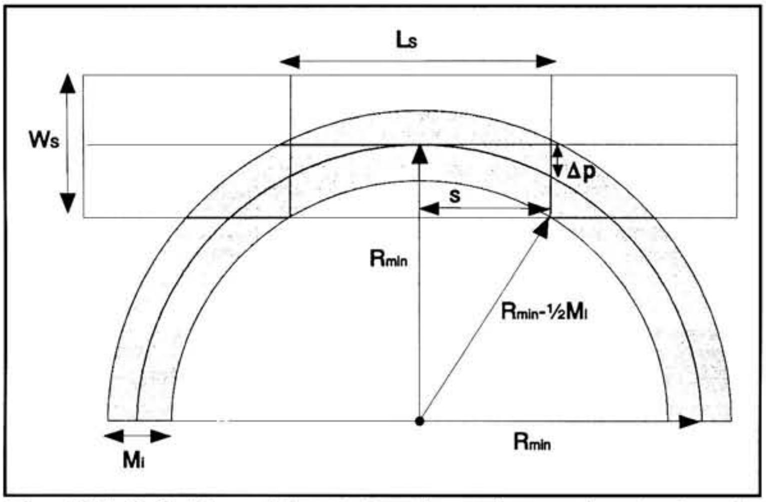

Fig. 5.34 Definition of length L_S of search area for circular slip road with radius R_{min}

5.7.2 EXAMPLES OF DETERMINATION OF PARAMETER SETTINGS

As an example of how the parameter settings in table 5.5 - 5.7 are determined, this is shown for two cases:

1) The width of the search area (W_S) and the threshold for the cross-correlation (R_C) of change detection, determined by a combination of theoretical derivations and experiments on aerial images
2) The angle difference ($\Delta\varphi$) between successive directions, determined by experiments on artificial images.

Example 1: combination of theoretical derivation and experiments on aerial images

There is a mathematical relation between two parameters of the image processing technique for change detection: the cross-correlation and the width of the search area. Their relation is derived in appendix C.1 for an ideal case. The threshold for the cross-correlation is derived from choosing the width of another road in the profile, which is considered significant for indicating a junction. The presence of a road that is more than one lane wider is chosen as threshold. Substitution of the standard width of a lane, Δw=3.5 m, and the width of 2-, 3- and 4-lane motorways respectively in formula c.5 yields the thresholds for the cross-correlation represented in the second column of table 5.8. These are maximum values, because due to noise in the real case, the optimal threshold should be lower. The factor which relates the width of the search area to the width of the road is set to 2 (see appendix C.1 for details).

For small sample sizes the cross-correlation coefficient R_C should exceed a minimum to obtain certainty that there exists similarity between the profile and the model. This means that the null hypothesis that there is no correlation, thus H_0: $R_C=0$, should be tested against the alternative hypothesis that $R_C>0$. A one-tailed student's t-test with N-2 degrees of freedom and a significance level α can be used to define the criterium for rejection of H_0, resulting in formula 5.1 [Hays, 1988].

$$\sqrt{\frac{N-2}{1-R_C^2}}\,|R_C| \geq t_{\alpha,N-2} \tag{5.1}$$

The number of pixels N in the profile corresponding to the minimum width of 2-, 3- and 4-lane motorways (table 5.2) at a resolution of 1.60 m and the t-test value for a 99% confidence interval yields the minimum values for the cross-correlation in table 5.8.

By experiments on the learning set of the aerial images the optimal threshold between the minimum and maximum is established. The threshold is increased from the minimum to the maximum with differences of 0.01. The detected changed parts of the motorways are drawn next to each other. The optimal threshold is determined visually for the images in the learning set considering the number of detected changed parts at positions where no junction is present and the number of changed parts belonging to more than one junction, which both should be minimal.

type of motorway	max R_C	min R_C	optimal R_C
2-lane	0.73 (w=11.55 m.)	0.65 (N=12)	0.70
3-lane	0.79 (w=15.20 m.)	0.55 (N=17)	0.75
4-lane	0.84 (w=20.75 m.)	0.47 (N=24)	0.79

***Table 5.8** Threshold R_C for cross-correlation*

Example 2: experiments on artificial images

The angle difference between successive directions is defined as the distance between the end points of two successive lines from the starting point. The main reason is that if the angle difference is defined in degrees, especially for large maximum distances to the search area, the distance between successive search area can be large, while for short maximum distances there is much overlap. The distance between successive search areas should be very small, because it should be prevented that none of the directions of the search area tested during rotation corresponds such to the direction of the link road that the value of the cross-correlation is larger than its threshold. If the angle difference is too large, the chance that these directions badly coincides is bigger, because there is an integer number of steps between the begin direction and the direction of the link road. This may result in accidental omission or detection. The influence on the cross-correlation is demonstrated in appendix C.2 in fig. C.9 and C.10 for an angle difference of 1.0 and 0.5 pixels respectively using an artificial image of a Y-junction (fig. C.7). It is shown that with an angle difference of 0.5 pixels the influence of this parameter on detection of Y-junctions reduces insignificantly.

5.8 DISCUSSION

One of the basics of knowledge-based interpretation is the separation of domain-dependent and object-specific knowledge from problem-solving computation [Hanson/Riseman, 1978], in order to make more clear which object properties are involved. These properties are compared with the properties of the segmented object during object recognition.
However, in this case study a part of the criteria for evaluation of segmented objects is defined within the image processing procedure. For example, only the first part of the link road with the best match with the artificial road model is returned based on evaluation of radiometric conditions. The reason is that our object definition only allows absolute evaluation of properties and not a relative evaluation which is needed to find the best match. But comparison of matches with all the variations in search direction, distance and model width can be done more efficiently within the image processing procedure than within the interpretation strategy. Because each image processing technique is used to search one specific object type, it is possible to see which properties are used to recognize each object type as parameters attached to the image processing techniques in the object relations.

The presented interpretation strategy and contents of the knowledge base are tested in the next chapter on the images of the learning and test set.