ABSTRACTION AND SCALE-SPACE EVENTS IN IMAGE UNDERSTANDING

Helmut Mayer
Chair for Photogrammetry and Remote Sensing
Technical University Munich
Arcisstr. 21, 80290 Munich, Germany
Phone: +49-89-2105-2688, Fax: +49-89-2809573
E-mail: helmut@photo.verm.tu-muenchen.de

Commission III, Working Group 3

KEY WORDS: Vision, Modeling, Artificial_Intelligence, Abstraction, Scale-Space

ABSTRACT

Image understanding can be described as the process of making information implicit in an image explicit in terms of objects. This implies a mapping of structured semantic information (symbols) to discrete noisy two-dimensional information (digital image). One way of solving this ill-posed problem is to fuse results in different images which have been produced from a single image by smoothing it with various degree. The smoothing reduces noise, but also changes the scale of the image, i.e. features are suppressed.
From a theoretical point of view the following questions arise: How can the abstraction of the description of objects be linked to the suppression of features in images of smaller scale? How can this be used for the recognition of objects?
As an answer to these questions, the link between abstraction and events in the so-called "scale-space" which mainly result in the elimination of substructures, is presented in this paper. Examples are sketched showing that this link has practical implications for the extraction of objects from images as well as for generalization of objects in geographic information systems (GIS) or cartography.

1 INTRODUCTION

Image understanding is a research area where, in spite of all big progress in theory and applications, its inherent complexity only slowly becomes clear. People think about the world in abstract concepts which describe a complex physical world. Concepts, like road, building, etc., are very hard to define in a formal way which can be implemented as a computer program. The central issue for a formal definition is knowledge representation (Sowa 1995). Another aspect is the knowledge which has to be represented. It has to come from an application (e.g. Photogrammetry, Remote Sensing, or GIS). Issues which have been addressed only recently are the importance of context (Strat 1995) or the combination of different kinds of information (information fusion (Clément et al. 1993, McKeown 1991)). A special case is the fusion of different scales (resolutions). That this can aid the interpretation has been shown in (Steger et al. 1995). Scale in the context of image sequences was treated in (Sester 1990). The goal of this paper is to give some theoretical considerations which support the use of multiple scales. Especially it will be shown how the semantics of objects is linked to scale, i.e. how scale can help to define concepts.

In psychology vision is thought of as a process that involves a lot of very specialized modules which interact in different directions (i.e. bottom-up and top-down). There was and is a quarrel between different researchers on the role of images, but newer results suggest that there really is a kind of image processing used in the path of reasoning. Kosslyn (1994) speculates that hypotheses generated on the line of reasoning about objects are verified by means of matching an image of the hypothesized objects created by computer-graphics like techniques into the real image.

Kosslyns findings furthermore suggest that distinctions are made between the description of singular objects and their spatial relations as well as between class and instance processing. Singular objects can further be used together with their spatial relations as substructure, i.e. parts, of more complex objects. Besides the fact that this constitutes a hierarchy of objects based on the part-of relation, there also is an abstraction linked to this. A settlement has for instance a substructure made of buildings, roads etc. But in addition to this it also has a new, more abstract, quality, as its own size or characteristics (shopping, recreation etc.).

Opposed to abstraction which deals with symbols, scale-space theory (Lindeberg 1994) is concerned with sub-symbolic signal (here: image) information. The scale-space is constructed by smoothing the original image with Gaussian kernels of successively increasing width. A property of scale-space theory is that additionally to the continuously evolving smoothing of the image events occur. These events are annihilation, merge, split and creation of extrema. Because most structure in an image, like points, edges, or lines are related to extrema, this means that also the structure is changed significantly.

An interesting question is, how these events are related to the abstraction of symbols describing objects in the image. Because most of the events will result in one way or another in the annihilation of extrema, structure will be lost, i.e. the information in the image is simplified. A light smoothing will mainly decrease noise. But greater amounts of smoothing will also destroy structure of objects (simplification). Substructure cannot be detected any more and the emphasis of the image is laid on the compound object. In certain ways this means that abstraction has occurred by simple smoothing.

The paper is organized as follows: In section 2 a short review on the term abstraction as well as a description of scale-space events is given. The link between abstraction and scale-space events is analyzed conceptually and empirically in section 3. In section 4 conclusions are given.

2 ABSTRACTION AND SCALE-SPACE EVENTS

2.1 Abstraction and Models

Looking at the term abstraction one finds that there are a lot of definitions for it. It can be defined as the "mental process of isolating a common element or explicating a relationship possessed by a number of things" (*Encyclopædia Britannica* 1985). According to Brachman (1983) abstraction is a relation of type "is-a" wherein a generic type is abstracted into an individual (e.g. "the eagle" in "the eagle is an endangered species"). In (*Encyclopedia*

of Artificial Intelligence 1985) the idea of abstraction is defined in the context of "search" as "to at first ignore the low-level details of the problem, concentrating on the essential features, and then fill in the details later." These examples show that the notion of abstraction is not generally clear.

In this paper a special notion of abstraction is used. It is defined in the context of image understanding where symbols are mapped to portions of images. The description by means of the symbols has to be structured. Additionally it has to be simplified, emphasis has to be laid on important things, and others have to be neglected. **Abstraction** is therefore defined in this paper as the increase of the degree of simplification and emphasis.

As has been pointed out in the introduction, this also has something to do with parts which construct the **substructure** of an object. Because, as Brachman (1979) states, the notion of a term has to be defined to enable a sound reasoning, the part-of relation is defined in this paper in terms of semantic networks following (Niemann et al. 1990) or (Mayer 1994). A **concept** consists of name, extension, and intension. The extension is the set of all objects which belong to the *concept*. The intension comprises all properties and relations an object needs to have to belong to a *concept*. Two *concepts* are linked by the **specialization** relation if the extension of one *concept* is a real subset of the extension of the other *concept*. The *specialization* relation defines an order among the *concepts*. More special *concepts* inherit the intension of more general ones. The **part-of** relation means the construction of a *concept* from other *concepts*. Representations, like the *concept*, or the *specialization* and the *part-of* relation, which are independent of an application are called epistemological primitives (Brachman 1979). In this paper other relations are used as well. But note that it is useful to restrict an actual implementation to the epistemological primitives, i.e. other relations have to be transformed to that primitives.

Simplification and emphasis are important characteristics of models (Rapp 1995) which are used to achieve the mapping of symbols and image data. This means that abstraction is also an implicit but integral part of models. And models are the critical basis of image understanding. They can be considered as the "theory" part of the theoretical framework of Marr (1982) as well as the conceptual level of the levels of knowledge representation of Brachman (1979). Explicit models have to be the foundation for every project in image understanding, because they can give reasons for deficits of an approach. Without an explicit model, i.e. if a system is only based on heuristics, no sound analysis of errors is possible and therefore the further development is hampered. The typical development will start with constructing a model from experience. The model is implemented and tested and according to the arising problems the model is improved. This is done iteratively.

2.2 Events in Scale-Space

Images are analog representations, representing non- or subsymbolic information by means of a homomorphism: The represented facts are contained in the representation. Relations between objects of the real world are transfered without loss of structure into relations of the representation.

The things which can be seen in an image are dependent on the scale (physical resolution). In a Landsat-TM image it is impossible to recognize a single human being on the ground whereas in an aerial image of scale 1 : 4000 this is easy.

Recently tools have been created for handling the concept **scale** in a formal manner. The main idea is the creation of a multi-scale representation by a one-parameter family of derived signals, where fine-scale information is successively suppressed (Lindeberg 1994). Data is systematically simplified and finer-scale details, i.e. high-frequency information is removed. The **scale parameter** $t \in \mathrm{R}_+$ is intended to describe the current level of scale.

The representation at coarser scales are given by a convolution of the given signal with Gaussian kernels of successively increasing width

$$L(x,t) = g(x,t) * f(x),$$

where $g : \mathrm{R} \times \mathrm{R}_+ \setminus \{0\} \rightarrow \mathrm{R}$ is the (one dimensional) Gaussian kernel

$$g(x.t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.$$

Another way to describe the evolution over scales is by means of a solution to the (one-dimensional) diffusion equation

$$\delta_t L = \frac{1}{2}\nabla^2 L = \frac{1}{2}\delta_{xx} L.$$

For the utilization of scale-space in discrete images a discrete scale-space theory has been developed (Lindeberg 1994).

One question which arises is, if it is not enough to carry out any kind of smoothing operation (e.g. mean). This is not the case because one of the features of smoothing of utmost importance is that in the transformation from the fine to the coarse scale no artifacts should be introduced, i.e. no new accidental structure should be created. Only the Gaussian kernel fulfills this criterion.

To describe the structure in an image, Lindeberg (1994) has defined so-called **blobs** as the (zero order) scale-space features. *Blobs* are closely linked to extrema in the image. They are smooth regions which are brighter or darker than the background and stand out from the surroundings.

In the process of smoothing the image there are four different discrete events which can happen to a *blob*: annihilation, merge, split, and creation. Whereas annihilation and creation are not too likely to occur (examples are given in (Lindeberg 1994)), merge and split of *blobs* are quite common. But *blobs* are only one means to represent the information content of an image. More commonly used representations are regions and edges (Haralick and Shapiro 1992). In a first approximation most of the events which can happen to *blobs* will happen to regions or their delimiting edges as well. In Figure 1 a) (see Figure 1 b) for thresholded versions of the normalized image) the image is gradually smoothed (from left to right; from top to bottom). The big region (upper left) is split into two regions (lower left) and these two regions are then merged again into one simple-shaped region (lower right). Other situations can be slightly more complicated. Imagine a "staircase" edge consisting of two edges connected by a small plateau between them. Edge extraction will result in two edges which are located close to each other. After smoothing only one edge will remain. Taking all this into account the term **scale-space event** is used for the remainder of this paper referring to events of regions and edges.

That annihilation is unlikely to occur only holds for ideal images. Figure 2 (original image) shows a car on the road, The image is gradually smoothed (from left to right; from top to bottom) until the car cannot be recognized any more. The level of smoothing where this happens depends on the level of noise in the image as well as on the closeness to other objects. Linked to this phenomenon are the **inner scale** and **outer scale** (Koenderink 1984). The *outer scale* is the (minimum) size of a window which completely contains the object while the *inner scale* is the scale at which substructures of an object begin to appear. For instance the car on the road can only be seen in the images in the upper half of Figure 2 (assuming good contrast *inner scale* corresponds to approximately 1m and *outer scale* to 4m resolution).

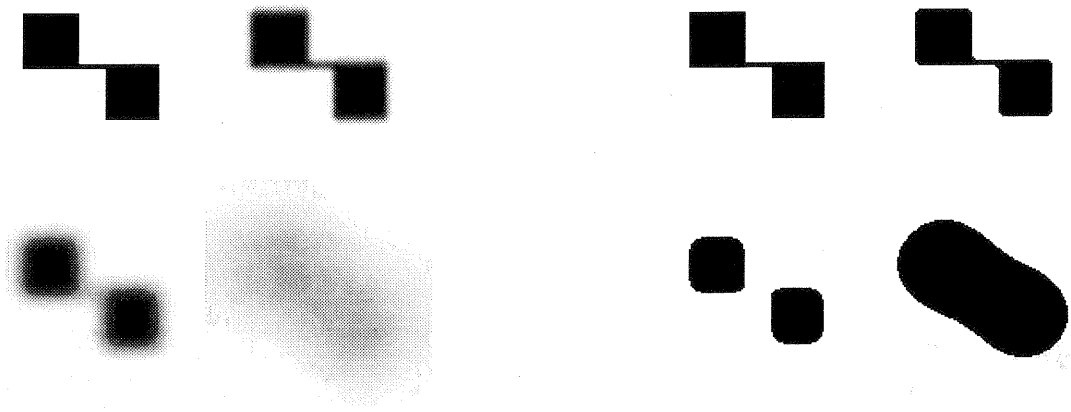

Figure 1: Events in scale-space — a) (left) Images b) (right) Thresholded images (from left to right from top to bottom: original image, $t = 2$, $t = 5$: split into two regions, $t = 20$: merge into one big region)

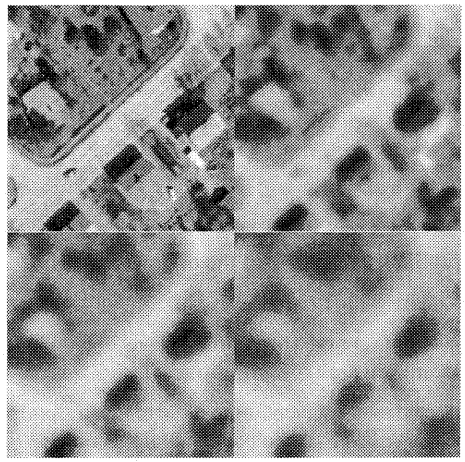

Figure 2: Car on road in scale-space — Images (from left to right, from top to bottom: original image, $t = 2$, $t = 3$, $t = 5$)

Figure 3: H — This is not what it seems to be

3 LINKING ABSTRACTION AND SCALE-SPACE EVENTS

One of the interesting properties of the human visual system is that it appears to represent information on multiple scales (Kosslyn 1994). Figure 3 shows information on two different levels of scale (similar patterns are used in psychological experiments). On a coarse scale you can see the letter "H", but on a small scale this is questioned ("this is not what it seems to be"). In this example the information on the two scales is independent but in many cases there is a close interaction of large scale and small scale.

How abstraction can occur by means of change of scale, how this is linked to *scale-space events*, and how this can be used in practical applications is shown in the following using the extraction of roads from aerial imagery and the generalization of buildings as examples.

3.1 Extraction of Roads from Aerial Imagery

There is a lot of work in the area of the extraction of roads from aerial and satellite imagery (Airault et al. 1994, Gruen et al. 1995, Jedynak and Rozé 1995, McKeown Jr. and Denlinger 1988, Steger et al. 1995). But only Steger et al. (1995) use different scales. This paper is in the line of Steger et al. (1995), but tries to add a theoretical foundation.

3.1.1 Large Scale: Roads in images of a resolution of 0.1m up to 1m pixel size, i.e. large scale, are complex objects (see Fig. 4 for a simplified semantic network representation). Roads, or more specific the road segments, are composed of the pavement (elongated region made of concrete or asphalt) with colored markings which are parallel to each other. Additionally, cars drive on roads, trees or buildings cast shadows on them or occlude them, and sidewalks lead parallel to them.

From this follows that roads can be detected either based on parallel edges which are the boundaries of the homogeneous elongated regions made of concrete or asphalt and the surroundings, or by directly detecting the regions by means of region growing. Figure 5 shows the result of region growing after selecting bright regions, eliminating small regions and closing of holes (see Fig. 5 b) for the original image). Even when using large thresholds to eliminate disturbances, as has been done here, parts of the roads where trees are casting shadows were not connected. A centerline representing the centers of the roads can be computed by simple thinning of the regions (Fig. 5 b)).

3.1.2 Small scale: In images of a lower resolution (2m up to 10m or more), i.e. smaller scale, cars on a road are not visible any more. The road segments are changed into linear objects which construct the road network (see Fig. 6 for a simplified semantic network representation). The road segments are connected by the crossroads. Special types of crossroads are simple crossroads or intersections. A specialization of a road segment is a bridge which crosses other objects.

Because a lot of detail is missing, the model represented by the semantic network has a higher degree of abstraction than that of the large scale. Roads appear as lines. Because of their

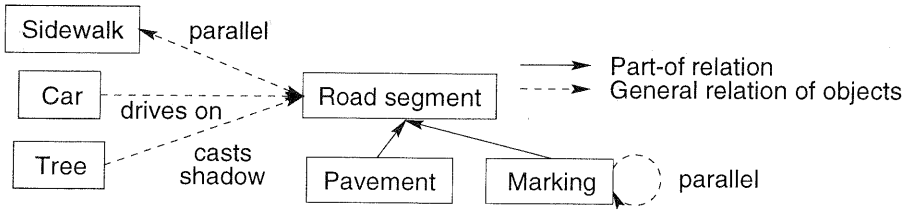

Figure 4: Model for large scale (concepts are depicted in boxes)

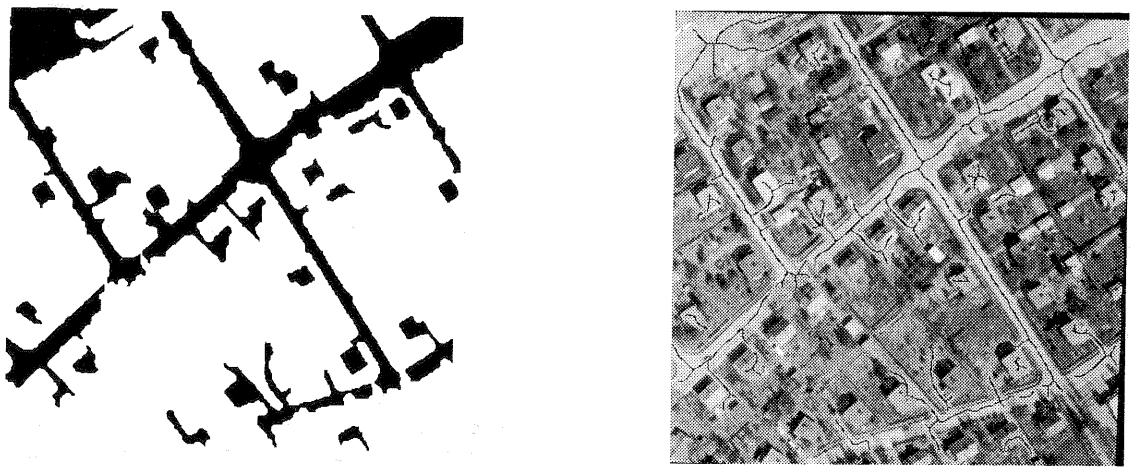

Figure 5: Region growing — a) (left) Large bright regions, b (right) Skeletons on top of the original image

material (concrete or asphalt), they mostly appear brighter than the background.

In Figure 7 hypotheses for centerlines of roads (bright lines) created in images of different resolutions (image pyramid; similar to scale-space but non-linear effects arise by subsampling) are shown. Lines are found using non-linear search, tracking and global non-maximum suppression with fixed parameters. Whereas in the image of larger scale (Fig. 7 a)) only some broken centerlines of the road from left to right have been found, the centerline was found quite stable in the image of smaller scale (Fig. 7 b); missing topological connections and the effects at the margin of the image are due to the implementation).

In the image of small scale it was evidently possible to detect the whole road network. Whereas in Fig. 7 a) the effects of the shadows have not been smoothed away sufficiently, they do not disturb the recognition of the lines any further (Fig. 7 b)).

3.1.3 Model: If one compares the large and the small scale, it is evident that information has been lost by eliminating regions and edges. This results in the elimination of noise as well as of substructure of objects. In Figure 8 the knowledge about road segments in the two scales is put together into one model. The model is split into three levels. The **real world** level consists of the objects and their relations on a natural language level. In the large scale a road is constructed of a pavement and the markings (solid or dashed) and cars drive or park on it.

The objects in the *real world* level are connected to the objects in the **geometry and material** level by means of the **concrete** relation (one of the other relation of section 2.1) which connects *concepts* describing the same object on different levels, i.e. from different points of view. The *geometry and material* level is an intermediate level which represents the three-dimensional shapes of objects as well as their material. This level has the advantage that it represents in contrast to the **image** level objects independent of sensor characteristics.

In small scale road segments are linked to mostly straight bright lines in the *image* level via the mostly straight concrete or asphalt lines in the *geometry and material* level.

In contrast to this the pavement of the large scale is linked to the elongated bright area in the *image* level by way of the elongated flat concrete of asphalt area in the *geometry and material* level. The markings are related to bright lines via colored lines and the car is a dark region as the concrete of a painted box made of glass and metal. Conceptually two things have happened here:

- The type of the geometrical representation has changed. The elongated area has been condensed into a line.
- But more important, the substructure of objects in the small scale (the markings, or the car on the road), has been eliminated.

This means that the complex object road segment composed of region-like pavement, and with markings and cars on it is changed into a more abstract linear road segment. Abstraction has occurred by means of elimination of substructure (annihilation) and merge of different parts separated by effects of noise.

3.2 Generalization of buildings

An area where multi-scale representations are closely linked to levels of abstraction are GIS and cartography. Whereas a map of very small scale contains only the largest countries and islands, a large scale map shows details of buildings. It is interesting that in maps the *outer scale* typically scales in proportion with the *inner scale* (Lindeberg 1994). This means that the content is adapted

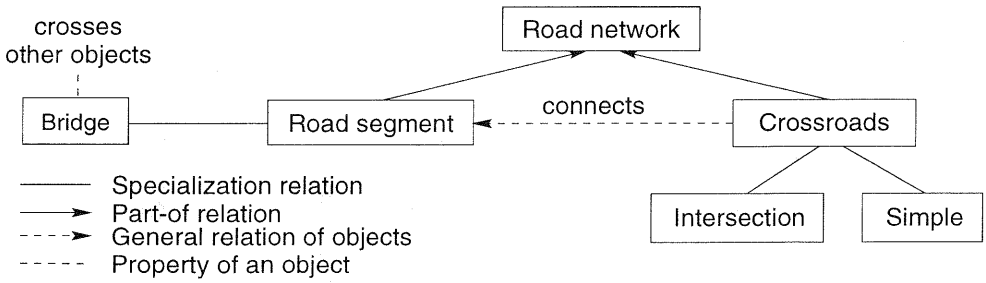

Figure 6: Model for small scale (concepts are depicted in boxes)

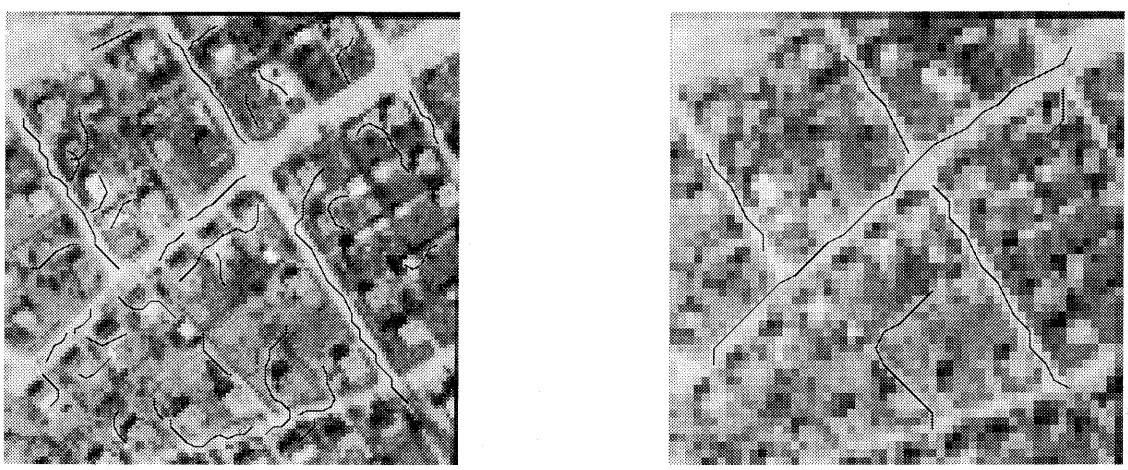

Figure 7: Detection of bright lines — a) (left) Larger scale, b (right) Small scale

to the physical resolution (scale) of the medium and the visual abilities of humans.

If digital data in a GIS is used, this adaption is often a problem. Figure 1 shows an example of effects of elimination of structure and simplification similar to cartographic generalization by means of *scale-space events* (see also section 2.2). For $t = 2$ the outline of the object is smoothed just a little bit. Then the small bridge between the buildings is deleted and two buildings are created by splitting ($t = 5$). After strong smoothing ($t = 20$) the two regions are merged again and one large simplified building is created.

4 CONCLUSIONS

Summing up, two things can happen simultaneously when an image is transformed by means of smoothing from a larger to a smaller scale:

- The information content of the image is reduced by eliminating regions and edges: Noise as well as of meaningful information are removed due to *scale-space events*.
- The removal of meaningful information when the *inner scale* is reached is synonymous with the elimination of substructure (parts) of objects and results in simplification. The interesting feature is that the loss of substructure often emphasizes the objects. According to the definition of *abstraction* given in this paper (increase of the level of simplification and emphasis) this is synonymous with an abstraction of objects.

Up to now, only Sester (1990) has treated the link between scale and abstraction of objects on a theoretical basis and there are few examples of a practical utilization. Opposed to the findings presented here, the goal of Sester (1990) was the analysis of image sequences and she did not emphasize the importance of representations on a different level of *abstraction* for the recognition of objects. Besides this, the concept of *scale space event* has not been developed at this time.

In conclusion, the outcome of this paper strongly recommends to use more than one scale for the recognition of objects. This has not only advantages in the performance but, as was shown, the emphasis put on an object by smoothing its substructure, i.e. using the *outer scale*, can be an important prerequisite for the recognition of objects ("only from a distance you can see clear"). A question which arises is the optimal scale to detect clues for an object. Small scales are especially suited for the detection of global structure and the formation of context while large scales add detailed/specific information for a detailed classification or verification of objects. Besides this, especially the complexity (form or material) of an object is important. In (Steger et al. 1995) a resolution of some meters was found to be useful for roads. A detailed analysis of this question for the objects of an application (e.g. road, building) could be, besides the question of context, one of the central issues of modeling, and henceforth performance, of image understanding systems in the future. This is one of the points where future research will be directed to.

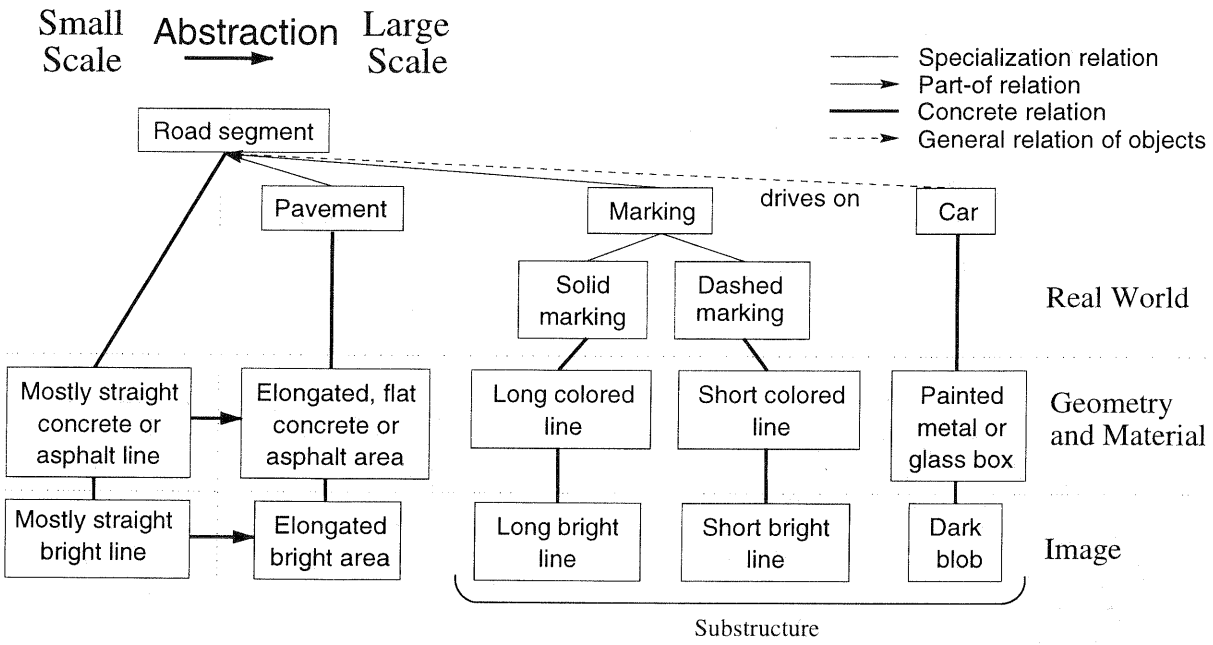

Figure 8: Model for road segment (concepts are depicted in boxes)

REFERENCES

Airault, S., Ruskoné, R. and Jamet, O., 1994. Road detection from aerial images: a cooperation between local and global methods. *in* J. Desachy (ed.), *Image and Signal Processing for Remote Sensing*, Proc. SPIE 2315, pp. 508–518.

Brachman, R. J., 1979. On the epistomological status of semantic networks. *in* Findler (ed.), *Associative Networks*, Academic Press, New York, USA, pp. 191–215.

Brachman, R. J., 1983. What is-a is and isn't: an analysis of taxonomic links in semantic networks. *Computer IEEE* 16(10), pp. 30–36.

Clément, V., Giraudon, G., Houzelle, S. and Sandakly, F., 1993. Interpretation of remotely sensed images in a context of multi sensor fusion. *IEEE Transactions on Geoscience and Remote Sensing* 31(4), pp. 779–791.

Encyclopædia Britannica, 1985. Encyclopædia Britannica, Inc.

Encyclopedia of Artificial Intelligence, 1985. Wiley & sons, Inc.

Gruen, A., Agouris, P. and Li, H., 1995. Linear feature extraction with dynamic programming and globally enforced least squares matching. *in* Gruen et al. (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser, Basel, Switzerland, pp. 83–94.

Haralick, R. M. and Shapiro, L. G., 1992. *Computer and Robot Vision*. Vol. I, Addison-Wesley Publishing Company, Reading, MA, USA.

Jedynak, B. and Rozé, J.-P., 1995. Tracking roads in satellite images by playing twenty questions. *in* Gruen et al. (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser, Basel, Switzerland, pp. 243–253.

Koenderink, J. J., 1984. The structure of images, *Biological Cybernetics* 50, pp. 363–370.

Kosslyn, S., 1994. *Image and Brain*. MIT Press, Cambridge, MA, USA.

Lindeberg, T., 1994. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Boston, MA, USA.

Marr, D., 1982. *Vision*. Freeman and Company, New York, USA.

Mayer, H., 1994. Automatic knowledege based extraction of objects of the real world from scanned maps. *International Archives of Photogrammetry and Remote Sensing* 30(3/2), pp. 547–554.

McKeown, D. M., 1991. Information fusion in cartographic feature extraction from aerial imagery. *in* Ebner et al. (eds), *Digital Photogrammtric Systems*, pp. 103–110.

McKeown Jr., D. M. and Denlinger, J. L., 1988. Cooperative methods for road tracking in aerial imagery. *Computer Vision and Pattern Recognition*, IEEE Computer Society Press, pp. 662–672.

Niemann, H., Sagerer, G. F., Schröder, S. and Kummert, F., 1990. ERNEST: A semantic network system for pattern understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(9), pp. 883–905.

Rapp, F., 1995. Modell und Realität. *Zeitschrift für Photogrammetrie and Fernerkundung* 6/95, pp. 220–223.

Sester, M., 1990. Multi-scale representation of knowledge based recognition and tracking of objects in image sequences. *International Archives of Photogrammetry and Remote Sensing* 28(3/2), pp. 868-888.

Sowa, J. F. (ed.), 1995. *Principles of Semantic Networks - Explorations in the Representation of Knowledge*. Morgan Kaufmann Publishers, San Mateo, CA, USA.

Steger, C., Glock, C., Eckstein, W., Mayer, H. and Radig, B., 1995. Model-based road extraction from images, *in* Gruen et al. (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhäuser, Basel, Switzerland, pp. 275–284.

Strat, T. M., 1995. Using context to control computer vision algorithms. *in* Gruen et al. (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser, Basel, Switzerland, pp. 3–12.