2007 Urban Remote Sensing Joint Event

Perceptual Grouping for Building Recognition in High-resolution SAR Images using the GESTALT-System

Eckart Michaelsen, Leo Doktorski
FGAN-FOM, Ettlingen, Germany,
mich@fom.fgan.de

Uwe Soergel[1], Uwe Stilla[2]
[1]Institute of Photogrammetry and GeoInformation,
University of Hannover, Germany
[2]Institute of Photogrammetry and Remote Sensing,
Technical University Munich, Germany

***Abstract*—GESTALT is a production system interpreter designed for advanced automatic recognition from difficult pictorial data. Building recognition from leading edge high resolution SAR-data is a good example for such a challenge. This contribution explains the system itself and its application to this particular issue. Perceptual grouping paradigms are coded in the productions in order to discriminate man-made ordered structure from arbitrary clutter. In particular symmetry and repetitive similar structure render promising prospects for this application.**

I. Introduction

Automatic building recognition from high-resolution SAR images still poses a hard challenge for contemporary image processing and extraction techniques. Human observers, however, even if they are not very familiar with the specific phenomena occurring in such data, will usually perform much better. We conjecture that specific perceptual grouping capabilities and hidden cognitive mechanisms may explain the stunning difference in the achievements between humans and contemporary machines in this field. On the other hand methods that are summarized with the term “cognitive vision” have recently regained much attention in the machine vision community [17] and there is considerable recent interest in structural methods [1] as well as in grouping techniques based on Gestalt psychology [2][5] again.

This contribution attempts to outline the cooperative and or competitive interaction of diverse inference modules in a structural cognitive vision system. We call the proposed architecture the GESTALT-system (Grouping Evidence System for Treatment of Alternatives within a Layered Task-solver). It emphasizes free combinatorial yet robust search for perceptual gestalts, while postponing decisions as far as can be afforded.

II. Related Work

A. *Traditional AI-Systems for Image Analysis*

Examples for traditional such reasoning systems have been the SIGMA-system [6] and the SCHEMA-system [3]. An interesting contemporary example for the combination of structural and statistic methods within a cognitive system where buildings are extracted from colored aerial imagery was published recently [13].

B. *Perceptual Grouping*

Local cooperation and conflict resolving in the context of perceptual grouping has been treated rigorously by the use of perceptual inference networks (PINs) [12].

III. Production Systems and Production Nets

Each **production** p is defined as quadruple $p=(\sum, \pi, \Lambda, \varphi)$ (input side $\sum$, constraint relation π, output side Λ, construction function φ). A finite set of such productions is called **production system**. Component names are short symbols. For visualization in production-nets we use one or two extra explaining words. For the meaning of productions the following possibilities are permissible:

- The components of $\sum$ are **parts of** Λ; in this case π gives the condition which the parts must fulfill in order to form the higher level gestalt. φ will construct higher level attributes such as symmetry axis, virtual contours etc.
- Some of the components of $\sum$ are interpreted as being a **context** such that the other components change there meaning accordingly and appear as such in Λ.
- The components of $\sum$ are interpreted as being a **concretization** of Λ. E.g. 3D-objects meant by oΛ appear as projected 2D-objects $\sum$ in the pictorial domain.

Part-of and context productions are present in the example system presented in Section III.B. A good example of a concretization production is given in [10]. There a production infers the presence and height of an elevated object from the layover displacements in its two appearances in two SAR images illuminated from different directions.

1-4244-0712-5/07/$20.00 ©2007 IEEE.

A. Compulsory Normal Forms

Originally ∑ and *Λ* are arbitrary n-tuples or finite sets of object symbols. However, without loss of generality they may be restricted to the following two gestalts:

$$(S_1, S_2) \xrightarrow[\varphi]{\pi} T \tag{1}$$

$$\{S, \ldots, S\} \xrightarrow[\varphi]{\pi} T \tag{2}$$

In both cases, π is a constraint formulated on the attributes of the components in ∑.”…” involved on the left hand side of the production form (2) denotes a set of components of arbitrary but finite size. The constructor function φ is defined with the object *T*. Note that in (1), the two left hand side components can be of different classes, whereas the left hand side components in (2) all must be of the same class.

B. The Example System

Up to now most systems consisted of only a hand full of productions. However, future systems will be much more complex. In this paper we will concentrate on the system given in (3). It codes Gestalt principles for perceptual grouping.

$$\begin{aligned}
&p_1 : \{L, \ldots, L\} \xrightarrow[\text{regression}]{\text{colinear}} LL \\
&p_2 : \{P, \ldots, P\} \xrightarrow[\text{average}]{\text{adjacent}} Sc \\
&p_3 : (LL, Sc) \xrightarrow[\text{initialize}]{\text{adjacent}} R \\
&p_4 : (R, SC) \xrightarrow[\text{append}]{\text{fitting}} R \\
&p_5 : (LL, LL) \xrightarrow[\text{inter sect}]{\text{adjacent}} A \\
&p_6 : (A, A) \xrightarrow[\text{axis}]{\text{symmetric}} Sy \\
&p_7 : \{Sy, \ldots, Sy\} \xrightarrow[\text{average}]{\text{adjacent}} C
\end{aligned} \tag{3}$$

This uses eight different types of components:

- ***P***-components are single image pixels that are saliently brighter than their surrounding. Instances of this object-class are constructed by feature extraction methods such as described in Section IV.A.2.
- ***Sc***-components are clusters of neighboring such ***P***-components. So they follow the perceptual grouping concept *proximity*.
- ***L***-components are short straight line segments. Instances of this object-class are constructed by feature extraction methods such as described in Section IV.A.2.
- ***LL***-components are prolonged straight line segments. These objects are too long to be extracted directly. They result from the perceptual grouping concept *good continuation*.
- ***A***-components are preferably rectangular angle objects. They are constructed from two ***LL***-components. *Rectangularity* is not among the classical perceptual grouping concepts. However it is useful for building recognition.
- ***R***-components are salient rows of bright scatterers. These objects are a very strong hint for building structures in high-resolution SAR images. They result from the perceptual grouping concepts *good continuation* and *similarity*.
- ***Sy***-components are symmetrically arranged pairs of ***A***-components. They result from the perceptual grouping concepts *symmetry*. This is known to be one of the strongest Gestalt principles of human perception.
- ***C***-components are clusters of consistent such ***Sy***-components. So they follow again the perceptual grouping concept *proximity*. Here however this is proximity in the space of symmetry axis.

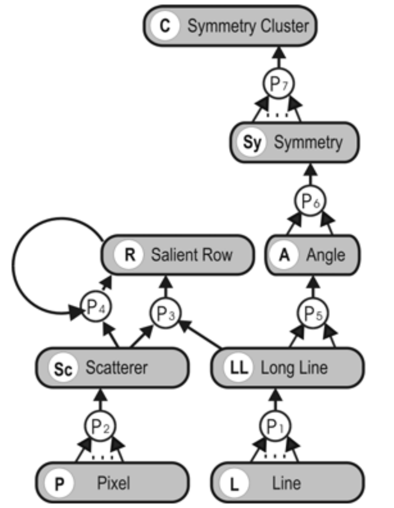


Figure 1. Visualization of the example System as production-net

In Fig. 1 the example system is presented as production net. Circular nodes represent the productions while elongated nodes represent component names. Such visualization helps the system designer understanding or explaining what the semantic connection between the objects is. In this respect it resembles a semantic network. However, it also indicates constraints on the workflow. In this respect it resembles a Petri-net: Connections are always flowing from a component-class to a production and from a production to a component-class. Interesting properties of this declarative representation can be seen at first glance such as the minimal serial depth, the presence of cycles etc. Also the combinatorial structure can be seen giving important hints for assessing the presumable computational complexity.

C. Accumulating Control

Listing all admissible productions for a given set of primitive components and given production system is not feasible in most cases. Instead, working-elements are formed for each component instance *c*. Such a working-elements has the form of a triple $h = (h_a, h_i, h_p)$ where $0 < h_a < 1$ is an assessment of *c*, h_i is an index or pointer allowing access to *c* and h_p denotes an index or pointer to a production (initially set to *nil*). Based on this notation, the following interpretation cycle can be formulated:

1) The working-elements are ordered according to the assessment h_a in a queue. The best *n* hypotheses are chosen. For each of these 2) is performed.

1-4244-0712-5/07/$20.00 ©2007 IEEE.

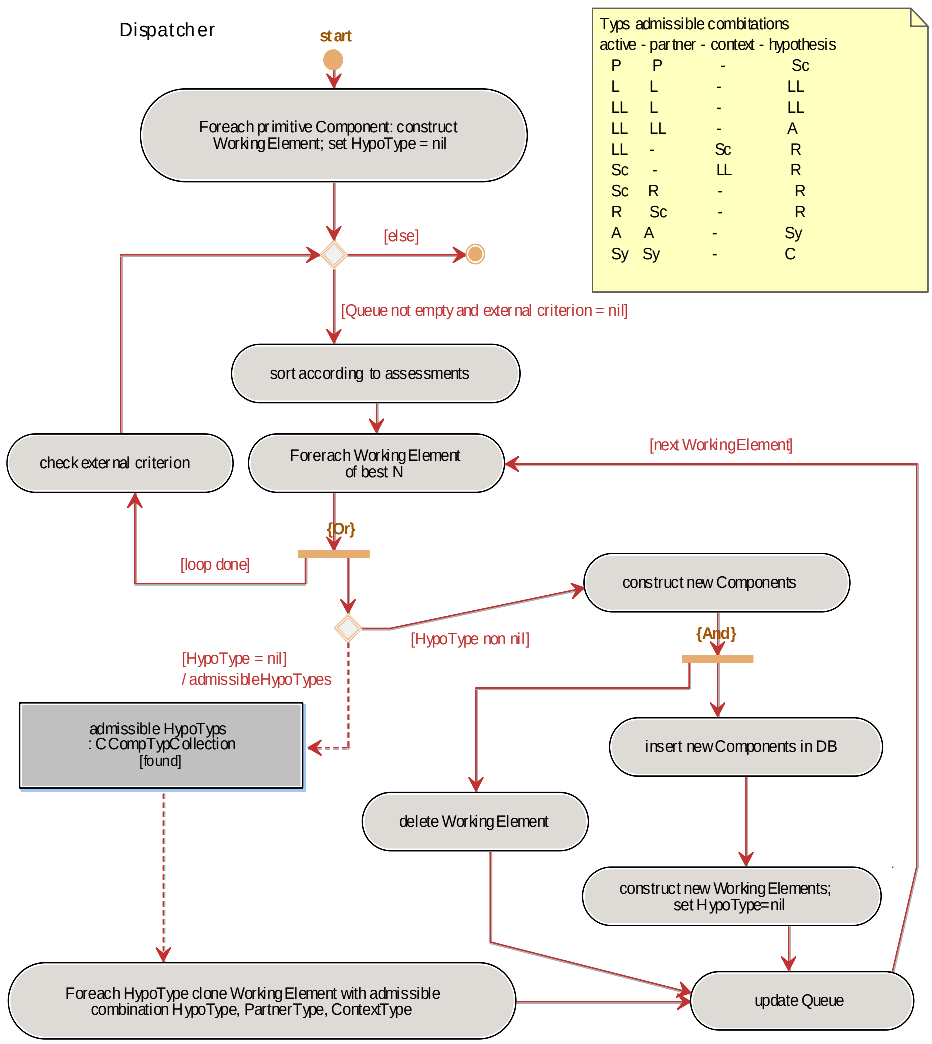


Figure 3. UML-state-diagram of the accumulating control scheme of GESTALT

2a) If $h_p = nil$ then all those productions are chosen, which contain an object of the same class as c on their left hand side. h is then replaced by clones of it with the corresponding indices h_p, where the assessment of those newly introduced hypotheses may be altered according to different importance of different productions. c is then called triggering object of h.

2b) Else there is a production to which h_p points with a corresponding right hand side object r. With this object there is a construction method φ which takes the triggering object c as fixed. It queries the actual set of components for the missing parts of the left hand side of the production h_p which fulfill the constraint π. Thus, the construction method φ may result in none, one or even many possible objects r. For all of these objects, new initial hypotheses are added to the queue. Thus – while the number of components is always rising – the number of working-elements in the queue may rise or fall.

3) While the queue is not empty, one may continue with step 1) or stop and return the current configuration as result. This result will be identical to the final unique solution, if the queue is empty. Otherwise, the current set of components will only resemble an approximate solution. An UML-state diagram of this interpretation cycle is given in Figure 3.

1-4244-0712-5/07/$20.00 ©2007 IEEE.

Figure 6 shows the graphical user interface that is displayed on the screen during the dispatch run. It shows a statistic on the current assessment values in the upper right section. Note, that this usually has a strong bias to badly assessed objects (below 0.2). This is intentionally made in order to prefer depth in the search early. Below the assessment curve the statistics of the current object numbers are displayed.

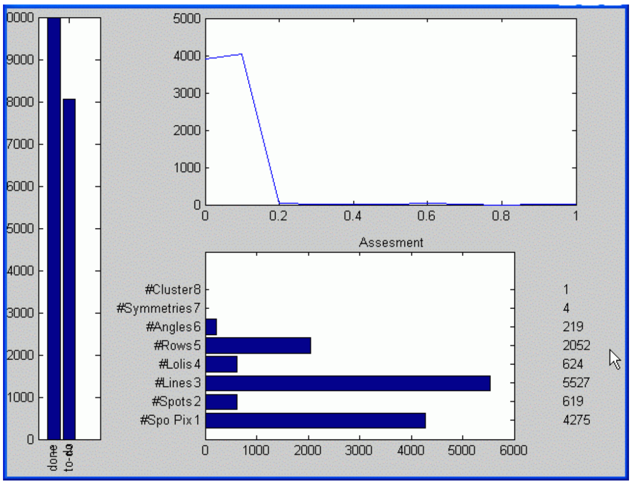


Figure 6. User interface of the GESTALT-system showing the current status of interpretation

IV. Experiments

Experiments with a collection of high-resolution SAR-images indicate robust and promising behavior of the proposed GESTALT system.

Figure 4 shows one of the images of a collection SAR images taken by the airborne PAMIR system of FGAN-FHR. In spot mode this airborne images can have even decimeter resolution. In order to apply the GESTALT-system primitive components have to be extracted from such pictorial data.

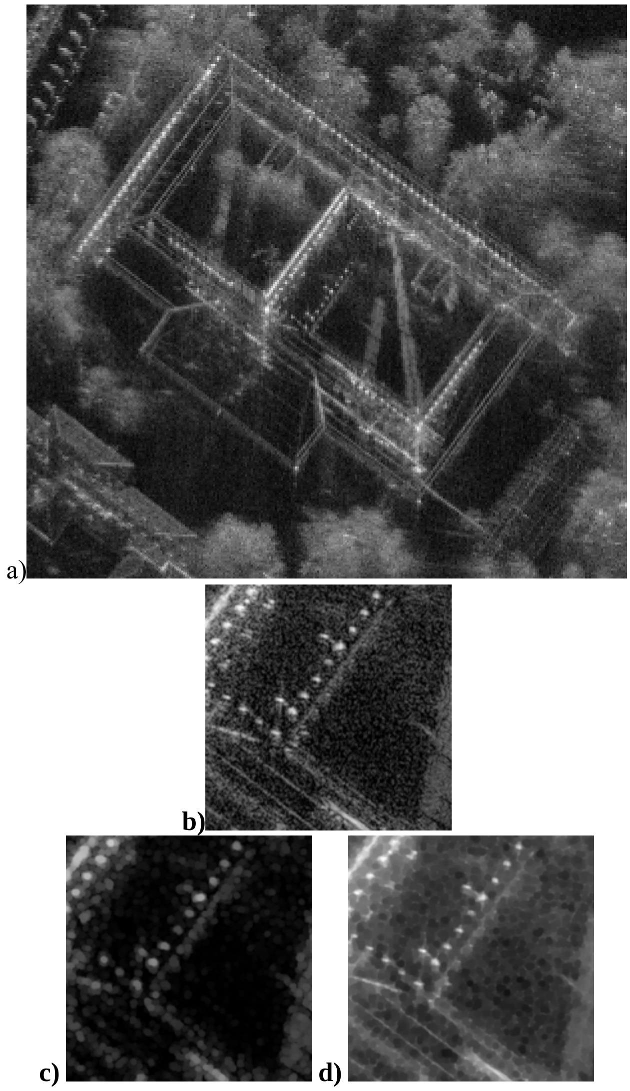


Figure 4. a) A high resolution SAR image; b) a section from it and filtered versions: c) opening d) closing

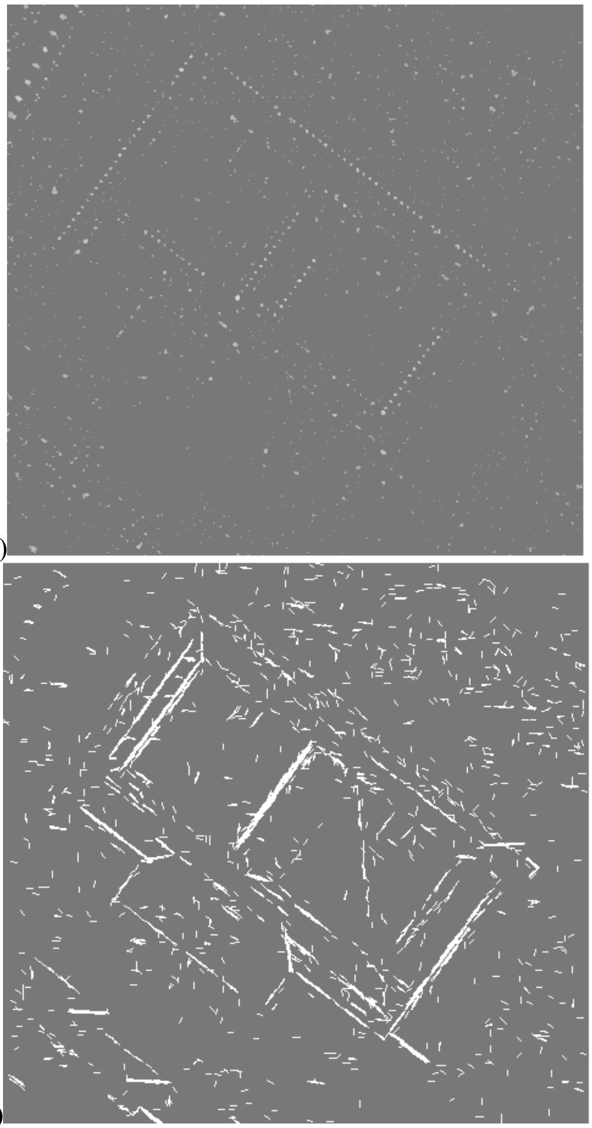
A. Extracting Primitive Components

A variety of techniques are applicable for the extraction from or the segmentation of the raw data. Usually it is also advisable to apply iconic filter operations in order to enhance the structures on which the analysis is based and to reduce noise which would otherwise unnecessarily burden the subsequent search.

*1) Iconic Preprocessing: Next t*o noise reduction some filters may also be used to enhance those features that are important for the task at hand. Among the many possibilities – such as LoG, diffusion filters or tensor voting [7] – morphological filtering is of particular interest for the kind of data processed here [16]. While most interest today concentrates on self dual filtering in order to treat bright and dark structure equally, for this particular task we know in advance that the building features to be extracted usually appear bright in SAR imagery. Therefore we can use morphological opening and closing. Figure 4 shows a small section of our example image in a) and the filtered versions in c) and d). Both reduce noise, but while opening almost perfectly isolates the bright spots, closing bridges the gaps and thus enhances the line structures. Bright spots or lines result from metallic or rectangular construction details on buildings that cause salient strong backscattering.

2) Spot Filtering: Object components ***P*** are extracted from the opened (and reduced) image by using a comparison filter and a threshold. Originally this filter has been deviced for Data from the thermal spectral domain [8]. However, it proved very usefull to extract salient bright scatters on high resolution SAR-imagery as well. Figure 8a shows the ***P***-components extracted from the image shown in Figures 4a) respectivly 4c).

3) Squared Averaged Gradient: Object components **L** are extracted from the closed (and tiled) image by using the squared averaged gradient [4] and a threshold. This filter requires three convolutions. It results in a 2x2 matrix for each pixel. Where the rank of this matrix is one, a line component is present. The resulting ***L***-components are given in Figure 8b.



a)

b)

Figure 8. Feature Extractiom: a) ***P***-components; b) ***L***-components

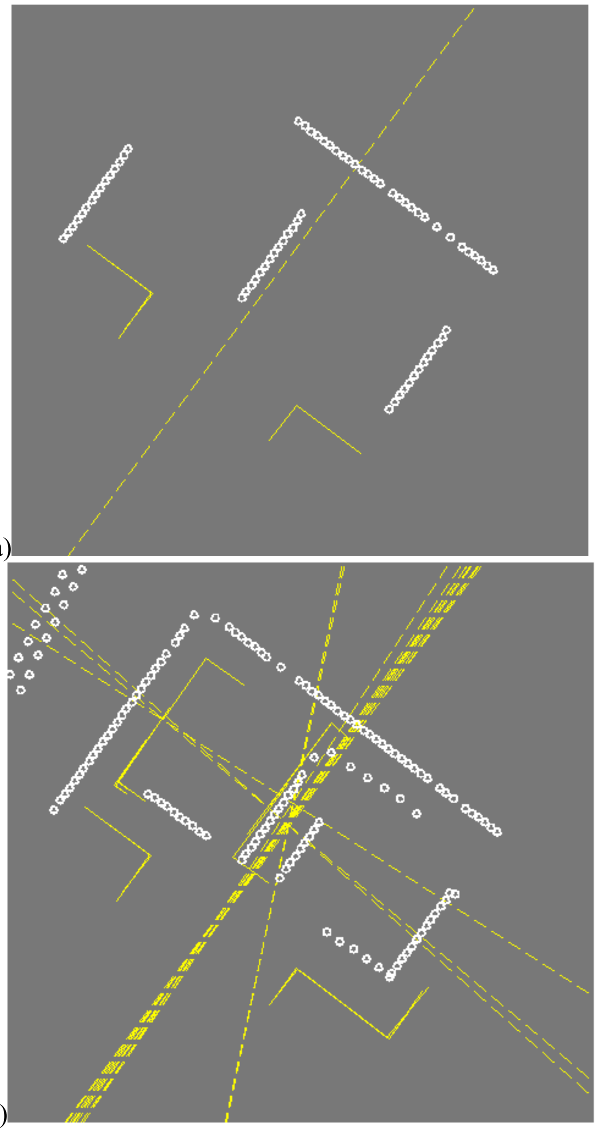
B. The systems behaviour in time

Once the primitives are extracted the interpretation cycle can be started. Table I shows the numbers of the different components as they evolve in time, while Figure 10 displays two states which were rather arbitrarily chosen. Symmetry cluster components are drawn in yellow and all salient row components with more than five members are given in white.

The major symmetry axis of the building is found rather early. And also the grouping of the salient scatterer rows starts at the desired locations. The second major symmetry axis appears a little later as more and more features of the building participate in the grouping process.

TABLE I. TEMPORAL DYNAMICS OF A GESTALT-SYSTEM

Cycles	Statistics of component numbers							
	P	**Sc**	**L**	**LL**	**R**	**A**	**Sy**	**C**
10.000	4275	619	5272	624	2052	219	4	1
20.000	4275	820	5272	881	4293	498	4	1
30.000	4275	820	5272	1031	7073	952	25	7
40.000	4275	820	5272	1118	10907	1068	41	14
50.000	4275	820	5272	1157	15366	1085	42	14



a)

b)

Figure 10. Result: a) after 10.000 cycles; b) after 50.000cycles

However, also false symmetry cues start to appear. These can probably be eliminated on a higher level reasoning state following the grouping process. Concerning the row components, the process on the structures of the central building seams to be saturated after 30.000 cycles. New rows are beginning to be grouped outside on other buildings.

V. Conclusion and Discussion

The presented experiments indicate robust and promising behavior of the proposed GESTALT system. However, there are numerous issues still to be addressed. The MATALB implementation on which this contribution is based is only an intermediate solution. It is too slow and still scales badly with rising numbers of component instances. A re-implementation of the system in object oriented and machine independent manner is under way. It is most important to include hashing mechanisms or other content-addressing techniques from data-banking into the productions in order speed up the search for potential partner components. In the preceding “blackboard” version of our work there used to be even special hardware for this purpose which was connected to the VAX-workstations [14]. The current implementation endeavor explicitly considers several possibilities for parallelization including data-banking mechanisms as well as consideration of FPGA-realizations.

Extremely important are graphical user interfaces for the parameter adjustment for the predicates and functions in the productions and also debugging tools which are constantly improved. There will also be a graphical user interface for the interactive design of new production-nets using a prefabricated pool of component definitions and productions.

Another important issue of ongoing research is the semantics of production-nets as used in the GESTALT-system. We see a trade-off between efficiency of the interpretation search and its formal soundness, which we see as prerequisite for such systems to have a meaning at all [11]. The dispatcher is modeled as inference machine for approximating meaning. This will require a result on convergence of the approximate solutions to the final true meaning. Practice – such as with this SAR-image interpreter – indicates, however, that solutions are best after some intermediate run-time and degrade later. So there also remains a lot of theory to be studied.

Repetitive structure, symmetry, good continuation etc. are very general concepts of human perception. Our contemporary automation proposal given above is quite preliminary and far too rigid. A real future cognitive perception machine should be able to recognize such properties everywhere – even if they occur unexpectedly.

Acknowledgment

We thank Prof. Dr. Ender and Dr. Brenner (both FGAN-FHR) for the high-resolution SAR data which were recorded by the PAMIR experimental system of FGAN.

References

[1] M. Basu, H. Bunke, A. Del Bimbo (eds): IEEE-PAMI Special Issue on Structural and Syntactical Methods, vol 27, no 7, 2005.

[2] A. Desolneux, L. Moisan, J.-M. Morel: *Gestalt Theory and Computer Vision.* In: Carsetti, A. (ed.) *Seeing, Thinking and Knowing*. Kluwer, Dordrecht, 2004, pp. 71-101.

[3] B. Draper, R. Collins, J. Brolio, A. Hanson, E. Riseman: *The Schema System*, IJCV, vol. 2, 1989, pp. 209-250.

[4] W. Foerstner: *A Framework for Low Level Feature Extraction*. In: Eklundh, J.-O. (ed). Computer Vision – ECCV 94. Vol. II, B1, pp. 383-394, 1994.

[5] D. W. Jacobs, M. Lindenbaum (eds): IEEE-PAMI Special Issue on Perceptual Organization in Computer Vision, vol 25, no 4, 2003.

[6] T. Matsuyama, V. S.-S. Hwang: *Sigma a Knowledge-based Image Understanding System,* Plenum Press, New York, 1990.

[7] G. Medioni, M. Lee, C. Tang: *A Computational Framework for Segmentation and Grou-ing*. Elsevier, Amsterdam, 2000.

[8] E. Michaelsen, U. Stilla: *Estimating Urban Activity on High-resolution Thermal Image Sequences Aided by Large Scale Vector Maps.* In: IEEE/ISPRRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Urban 2001, IEEE, pp. 25-29, 2001.

[9] E. Michaelsen, U. Stilla: *Probabilistic Decisions in Production Nets: An Example from Vehicle Recognition.* In: T. Caelli, A. Amin, R. P. W. Duin, M. Kamel, D. de Ridder (eds): *Structural, Syntactic and Statistical Pattern Recognition SSPR/SPR 2002,* LNCS 2396, Springer, Berlin, pp. 225-233, 2002.

[10] E. Michaelsen, U. Soergel, U. Thoennessen, *Potential of Building Extraction from multi-aspect high-resolution Amplitude SAR-Data*. Joint workshop of ISPRS/DAGM Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring CMRT, International Archives of Photogrammetry and Remote Sensing, vol. 36,part 3/W24, pp. 149-154 .2005.

[11] E. Michaelsen, M. Arens, L. Doktorski, P. Ding, U. Stilla: *Meaning and Efficiency in the GESTALT-System.* Accepted for VISAPP07, 2007.

[12] K. Sarkar, K. L. Boyer: *Computing Perceptual Organization in Computer Vision*. World Scientific, Singapore, 1994.

[13] H.-F. Schuster: *Detection of Man-made-objects Based on Spatial Aggregations,* http://www.ipb.uni-bonn.de/ipb/projects/OntoSkalen/html/TB-ipb-05_1.pdf, 2005.

[14] U. Stilla, E, Michaelsen, K. Lütjen: *Structural 3D-analysis of aerial images with a blackboard-based production system.* In: A. Grün, O. Kübler (eds): *Automatic extraction of man-made objects from aerial and space images*. Birkhäuser, Basel, 53-62, 1995.

[15] U. Stilla, E. Michaelsen, U. Soergel: *Perceptual grouping of regular structures for automatic detection of man-made objects - Examples from IR and SAR,* IGARSS 2003.

[16] P. Soille, M. Pesaresi: *Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing,* IEEE Trans. on Geoscience and Remote Sensing, vol. 40,no. 9, pp. 2042-2055, 2002.

[17] D. Vernon: *Cognitive Vision – The Development of a Discipline*. http://europa.eu.int/information_society/istevent/2004/cf/document.cfm?doc_id=568, 2004.