A CONCEPTION FOR OBJECT-ORIENTED DESCRIPTION OF KNOWLEDGE* AND DATA IN SCENE ANALYSIS AND OBJECT RECOGNITION

Yonglong Xu, Msc., Institute of Photogrammetry and Engineering Surveys,
University of Hannover, Nienburgerstr. 1, 3000 Hannover 1, BRD
Presented Paper to Commission III, ISPRS

Abstract

A conception for object-oriented description of knowledge and data is suggested, aiming at semiautomatic or full-automatic image analysis and object recognition and integration of the results into a GIS. The conception itself can also be treated as an integrated GIS model.
An object is defined by its attributes, method or action slots and relations with other objects. The world is represented in the knowledge base by 2 object nets: a class hierarchy for semantics, a membership hierarchy for physical objects and geometric data which are stored in either raster or vector form. Photogrammetric data are also represented in the object form and stored in the data base. The object recognition is carried out by finding out instances of class objects from the data base and filling the membership hierarchy with them. Some preliminary results are reported, e.g. structures for knowledge and data description, mutual transformation of external and internal representations, the work in the low-level processing and middle-level processing and some useful algorithms. Several open problems e.g. knowledge acquisition, reasoning for object interpretation and resegmentation are discussed.

KEY WORDS: object, object-oriented description, knowledge base, image analysis, object recognition.

1. INTRODUCTION

The development of photogrammetry and remote sensing is today characterised by more and more handling various kinds of digital information with computer-aided or automatic approaches. For efficient use and management of such information Photogrammetry and remote sensing should be integrated with geographic information systems(GIS), where their main task is the collection and updating of data for GIS from remotely sensed data/Willkomm, 1991/. This tendency is especially reflected on the recent EARSeL Workshop on Relationship of Remote Sensing and Geographic Information Systems. Although several interactive data acquisition systems are already developed by different companies, such as PHOCUS from the company ZEISS /Willkomm, 1991/, automatic image interpretation and identification by using image understanding techniques remain to be a problem /Li, 1991/.

Since 1960's it has been tried to recognize objects on images. It's realized, that it's almost impossible to interpret complicated objects on images automatically, just using general image processing routines without introducing semantics and other background knowledge, which are usually owned and governed by human experts. With this thought Artificial Intelligence (AI) and Expert Systems (ES) were created as a new science to treat this kind of problems. At the beginning some initiators in this field were very optimistical, believing that really intelligent machines which could think and see could be invented in the near future/Zhang, 1986, 1987/. But the practice has proved that the tasks are extremely hard and so far the progress has not been so great as expected. Nevertheless this hasn't disturbed further relevant researches in this science and reasonable use of ES techniques in different fields, considering that the above thought represents the correct guide direction. Instead a lot of Efforts have been made to apply AI and ES to solving problems, where human expertise is needed, not only in other scientific fields, but also in image analysis and understanding, for example in medicine/Niemann, 1987; Towers, 1988; Vernazza, 1987/, photogrammetry and remote sensing/Brooks, 1983; Compbell, 1984; Lambird, 1984; Mooneyhan, 1983; McKeown, 1984; Nagao, 1980; Riekert, 1990/, and other industral applications/Liedtke, 1989; Pentland, 1986/. Many interesting results have been reached, which show the good promise of these techniques and approaches.

The first but decisive step for an ES is to design a suitable information representation structure. This is especially true for a vision ES. Considering the supposition that any conceptual or physical object can be approximated through a structural reduction of it into a set of simple but with each other associated elementary parts /Brooks, 1983; Fu, 1987; Pan, 1990/, such a structure should be supported by facilities for data consistency, effective data updating, multi-inheritance, various

* The work, supported financially by the Gottlieb Daimler and Karl Benz Foundation, was made under the supervision of Prof. Dr. mult. G. Konecny.

data types and description of multirelations and interface to external environment. Not all available object-oriented systems or ES-shells fulfil all the requirements. This motivate us to develop our own system for object-oriented description of knowledge and data. It aims at semiautomatic or full-auotomatic analysis and recognition of objects on aerial images and integration of the results into GIS. An object is defined by its attributes, method or action slots and relations with other objects. The world knowledge is represented in the knowledge base (KB) by two object nets: a class hierarchy for object semantics, a membership hierarchy for physical objects and geometric data which are stored either in raster form or in vector form. Photogrammetric data are also represented in the object form and stored in the data base(DB). Objects on images are interpreted by building the connection between the KB and DB, determining instances of class objects from the DB and filling the membership hierarchy with them. The characteristics of the system are that any data are described by an uniform structure i.e. object; each object can have fathers of any number with different father-son relations and the level depth of each hierarchy has also no bounds, so long as there is still memory space. Dynamic updating and interface to C-functions and UNIX are embedded. This structure itself can be used as a integrated GIS model. The communication with an existing GIS can also be erected by a corresponding interface.

In this paper the solution of image understanding using AI and ES will be briefly reviewed at first. Then the model of the object-oriented description will be presented, where its internal and external structures, characteristics and advantages will be explained. Afterwards the up to now obtained results and other necessarry preprocessing proceedures are concisely stated, e.g. mutual transformation of external and internal representations, image preprocessing, segmentation and its description. Finally several open problems e.g. knowledge acquisition, reasoning for object interpretation and resegmentation are mentioned, which are to be studied and worked out in future.

2. SOLUTION OF IMAGE UNDERSTANDING USING AI AND ES

Image understanding is just to recognize and describe concerned objects from input grey-level or iconic images. In general, the solution using ES consists of three phases in different levels (Fig. 2.1): in the low level processing(LLP) iconic images are first divided into in some sense meaningful segments - for the purpose here region-oriented segmentation methods are normally used; the segments (regions) are then described as primitives according to thier attributes and relations in the middle level processing(MLP) - all of these are main part of the DB; in the high level processing(HLP), with the support of DB and the information in the KB, which is in advance acquired as the model knowledge for the interested objects, objects on images are finally interpreted by the reasoning mechanism through setting up the connection between the KB and DB and filling the membership hierarchy with the instances of the class hierarchy found in DB. As the result a more complete description of the objects is obtained. The system architecture in HLP is shown in Fig. 2.2, which has four components: reasoning machine, KB, DB and interface to GIS.

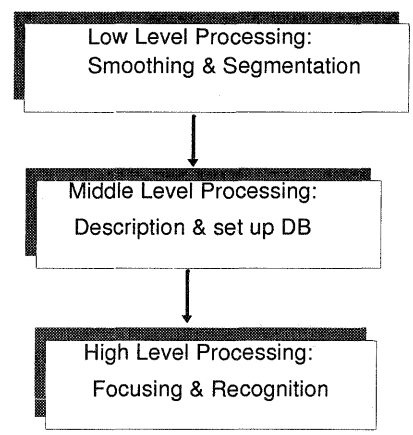


Fig. 2.1 solution for image analysis using ES

To date, understanding primitive perception as the interpretation of sensory data by use of models(konwledge) of the world is the standard vision research paradigm /Pentland, 1986/. Since objects in the real world are modelled, a system built up on the above scheme is also model-based. Such a system is so called blackboard one, when special state areas are opend for recording processing states, which are modified by system actions, and only change in states can cause and start new actions of the system. If the actions are independent of each other and data-driven, the system is also named production system/Negoita, 1985/. The great advantage of such a system is thier distinguished modularity.

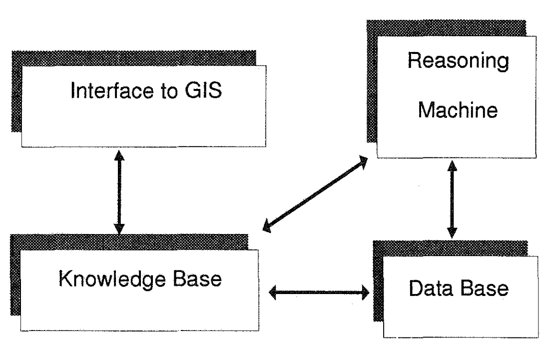


Fig. 2.2 knowledge based image analysis in HLP

3. THE MODEL OF THE OBJECT-ORIENTED DESCRIPTION

3.1 general explanation

As a basic knowledge representation unit an object is made up of several slots and each slot can have several facets(Fig. 3.1). Depending on concrete requirements, slots can be filled with different attributes, methods, actions and relations with other objects and facets can be filled with corresponding values. In the literature objects with this structure are sometimes called frames or units /Bauer, 1987/. The internal structures of such descriptions are actually semantic nets.

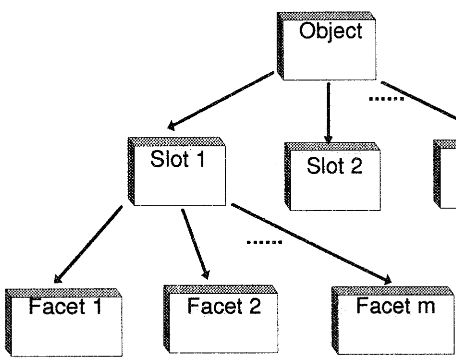


Fig. 3.1 structure of an object

This section deals with structuring the KB and DB implicitly and explicitly with the object defined above as uniform elementary building piece. In this way the KB and DB consist of a collection of objects respectively, with the features of readability, modifiability and extensibility. This object-oriented description was so designed and implemented that it is characterized by its inheritability, which is enabled by relation IS_A, and no limitation to father number and level depth, which is by all means necessary because of the complexity of scene objects on the ground and makes 3-D object description possible.

3.2 Internal implicit structure

The KB consists of two semantic nets: the class hierarchy and membership hierarchy. The DB is a semantic net. Each net is composed of a set of objects. A table of global parameters is used to put some common parameters or specify the semantic nets, e.g. the sizes, total number of objects, total number of attributes and so on.

The class hierarchy represents general semantic knowledge about concept or class objects in analysis, which looks like the following:

-- level 0 : root, i.e. world model - scene;
--- level 1 : concept or class objects, e.g. buildings, waters, roads, etc.;
--- level 2 : 1. subconcept or subclass objects, e.g. streets, railways, etc.;
.
--- level n : (n-1). subconcept or subclass objects.

The relation between any 2 levels is IS_A.

The membership hierarchy, which is as follows structured, represents concrete physical objects in scene and contains specific local knowledge about them. Therefore it can be a digital map model.

--- level 1 : physical objects, e.g. objects on digital maps;
.
.
.
--- level m : primitives: regions in a compressed raster-form.

The relation of level 1 with level n in the class hierarchy is IS_A. And the relation of level m with level m-1 is PART_OF. A compression method with quick retrieval ability is developed for representing primitives in raster form. However they can also take vector data form, which are not supported at present.

Photogrammetric raster data are stored in the DB, which has the following structure:

--- level 1 : primitives: segmented regions;
--- level 2 : a label image;
--- level 3 : a quadtree description of input grey images.

Each object(region) in level 1 has several attribute slots, of which one represents the recognition state of the region. Level 1 also takes the compressed raster form, so that superposition manipulation with level m is enabled after geocoding.

Besides IS_A and PART_OF, which were introduced in the implementation as son-father relations to erect the hierarchical connection between objects in two different levels, relations like INSTANCE_OF, MODEL_OF, CONTAINED and NEIGHBOR_OF are also available. Other topologic and geometric relations, however, are also allowed to be built into this description to form new son-father relations, with which, together with the above relations, a complete 2-D or 3-D object description can be only structured.

The above structures are supported by function libraries for object creation and storage, data accessing, attribute inheritance, dynamic fast updating, data consistency and interface to external environment.

3.3 External explicit structure

The external structure of the object-oriented description is concretized by the external definition of an individual object. In our implementation an object and its attributes are defined currently as follows:

```
    OBJECT                        string-name
1) *INSTANCE-NUMBER               number   (default: 0)
   *FATHER-OBJECT                 object1, object2, ...
   *RELATION-WITH-FATHER string1, string2, ...
                                  (default: IS_A)
    SON-OBJECT                    object1, object2, ..
                                  (can be omitted)
    LEVEL-IN-TREE       number    (can be omitted)
    ATTRIBUTE-NAME                attribute1, attribute2, ...
                                  (can be omitted)
    END

    ATTRIBUTE                object . attribute
1) *VALUE-NUMBER             number       (default: 0)
   *VALUE-TYPE               INT, STRING, FLOAT,
                             BITFIELD  (default: INT)
   *INITIAL-VALUE            value1, value2,...
   *DYNAMIC-VALUE            C-function or
                             external program
    END
```

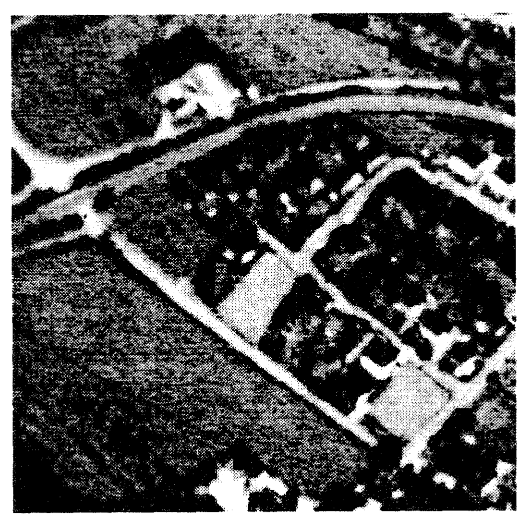
The slots marked with * can be omitted or empty, if they don't exist. The slot names can be redefined by the user himself and the sequence of the slots is indifferent. The slots marked with 1) can be filled with variables (object . attribute or attribute), C-functions or external programs, which are automatically distinguished. This mechanism is very important for data consistency. Besides VALUE-TYPE can be filled with SET-INT, SET-STRING, SET-FLOAT, RASTER-CHAR and RASTER-INT. The syntax isn't strict but more or less fuzzy and natural, e.g. as seperator between items you can use space, ',' or ';'. The above definitions are fundamental and will be extended and modified further to fulfill new requirements.

4. FINISHED PREPARATION WORK

As a main part of HLP the whole conception has been realized in language C on a SUN-workstation. For the actual knowledge based HLP the to be interpreted multispectral images are prepared in LLP and MLP. In order to get a satisfactory segmentation the images are first smoothed with the edge-preserving method proposed by /Nagao, 1980/, so that the regions become as homogeous as possible without destroying edges. the smoothed images are segmented with the quadtree-method developed by the author, where it is expected that the backtracking in HLP can be eased/Xu, 1991b/. Fig. 4.1 and 4.2 illustrate the smoothing and segmentation of an image respectively. Afterwards in MLP the DB is automatically generated with several basic attributes for the segmented primitive regions, which are for instance encasing rectangle, average grey level and area of a region, etc.

For HLP four important algorithms have already been developed:

- fast determination of region neighborhood and containing relations.
- fast raster-vector and vector-raster transformation for hybrid and simultaneous processing of raster and vector data and for manipulation of individual objects /Xu, 1991a/. If original data are in raster-form, the transformation and back transformation are 100% identical. Thus the geometric data represented in raster-form are prefered here.
- fast heuristic detection of region corners for shape analysis. An example for several regions in Fig. 4.2 is shown in Fig. 4.3, in which the regions are twice enlarged for better examination.
- mutual transformation between external and internal knowledge representations.



a. result of a histogram equalization

b. result of an edge-preserving smoothing

Fig. 4.1 a cutout from an aerial image

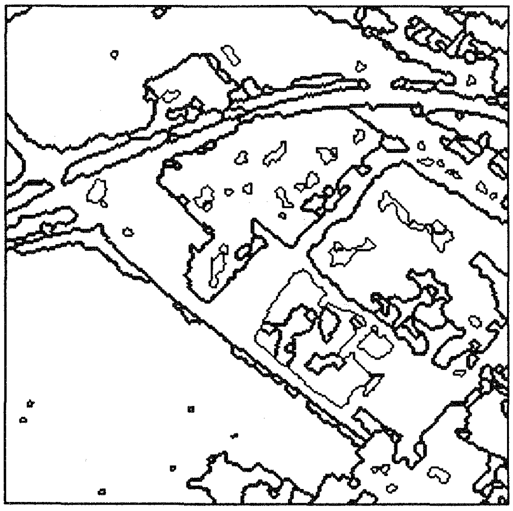


Fig. 4.2 segmentation of the cutout

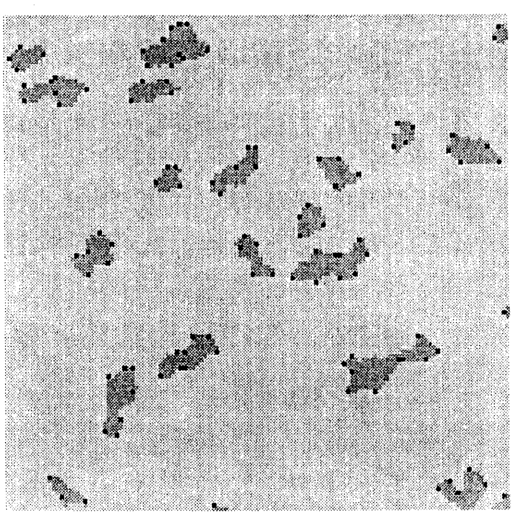


Fig. 4.3 determined corners of several regions in Fig. 4.2

5. DISCUSSION

With emphasis on the knowledge representation, we have presented our idea to develop a knowledge based system for scene analysis and unterstanding in the above sections. Still we can't expect that machine vision problem can be solved without great difficulties, because what ES and AI methods offer us are nothing else but programming tools. Nevertheless, we can better modularize knowledge and reasoning by using these methods, so that they can be kept easy to read and modify, and easy to take distributed and parallel processing. The above object-oriented description makes local and structural image analysis possible, which is highly important because of limitations of global and statistical analysis. Two crucial and most difficult problems remain unsloved: a learning mechanism for acquiring knowledge and transforming it into internal representation; a reasoning mechanism for finding out and correcting segmentation errors and for recognizing objects.

In this context the goal of learning is to fill the KB with suitably structured knowledge about objects automatically or semiautomatically. A deductive learning mechanism is already enabled by the suggested object-oriented description: description structures of individual objects are first analysed and worked out, afterwards are generated interactively as external descriptions, which are finally converted into internal forms by the program for external-internal transformation. The KB can also be set up inductively (i.e.example-driven): the membership hierarchy is first gained through an available interface to GIS, with which the class hierarchy is then richened and extended, e.g. adding new object species. The same can be done by using interpretation results from the DB or both the results and the membership hierarchy. For human examination and check an internal-external transformation can be executed.

The reasoning for object recognition is realized in three steps, which should be supported by a number of algorithms for geometric and relational object analysis in different levels:

- focusing: preclassify regions in the DB to find candidates for certain objects in order to improve the efficiency.

- backtracking through resegmentation: detect segmentation errors and correct them. A complete and perfect segmentation would greatly simplify the followed HLP. Unfortunately for remotely sensed data it has proved to be tremendously difficult and even impossible owing to insufficient information during segmentation and the difference between human perception and mathematical algorithms. For this problem there exist two solving methods. One of them emphasizes that semantics should be introduced into segmentation so as to guide and improve it, while in the other segmentation errors are expected to be found out in HLP and then corrected by backtracking. The former integrates the bottom-up and top-down strategies together and is restricted to certain applications owing to the early introduction of semantics; the latter uses first the bottom-up and then the top-down strategy and is adopted here.

- object recognition through reasoning: about handling knowledge uncertainty there are two reasoning strategies: inexact reasoning and approximate reasoning which is based upon fuzzy-set theory and mathematically sound. In many model-based vision systems the problem is indirectly handled by a more or less exact reasoning, where uncertainty is implicitly contained in the knowledge body and reasoning. The difficulty of a explicit treating lies in determining a suitable certainty measure for facts and rules. For efficient verification of a fact backward chaining is very helpful: a hypothesis is first generated with help of available information and then verified. In addition, the reasoning should be possible with and without the membership hierarchy.

In addition to the above to be investigated bottleneck problems several other technical problems need to be studied yet, before a complete workable system can be created:

- extend the description structure to handle more spatial or topologic relations, in order to model various objects in 2-D or 3-D space.

- establish the interface to GIS, although the conception itself can also be considered as a integrated GIS model.

- explain results by using information in the KB and DB.

- seperate implicit domain-dependent knowledge(procedures or rules) from reasoning operations, formulte it modularly in terms of individual functions, so that a blackboard system can be completed.

- handle huge data amount through skilled programming, e.g. using a bit for representation of states, storage space mapping and allocation techniques etc.

REFERENCES

Bauer, k.; I. Büttel; L. Eberhard; M. Hälker; H. Lehner; K. Micholka und R. Paulsburg (1987) : Expertensystem - Einführung in Technik und Anwendung, Siemens Aktiengesellschaft, Berlin und München, 1987.

Brooks, R. A. (1983): Model-Based Three-Dimensional Interpretations of Two-dimensional Images, IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No.2, March 1983.

Compbell, W. and Roelofs, L. (1984): Artificial Intelligence Applications oncepts for the Remote Sensing and Earth Science Community, the roceedings of Spatial Information Technologies for Remote Sensing oday and Tomorrow, IEEE Computer Society, USA, 1984.

Fu, Kingsun; Z. Cai and G. Xu (1987): Artificial Intelligence and its Application, Publishing House in Qinghua University, Beijing, 9.1987.

Lambird, B. A.; D. Lavine and L. N. Kanal (1984): Distributed Architecture and Parallel non-Directional Search for Knowledge-based Cartographic Feature Extraction Systems, Int. J. Man-Machine studies (1984), 20, 07- 120, 1984, Academic Press Inc. (London) limited.

Li, Deren (1991): From Photogrammetry to Iconic Informatics - on the Historical Development of Photogrammetry and Remote Sensing, Department of Photogrammetry and Remote Sensing, WTUSM, Wuhan, 1991.

Li, Deren and Jianya Gong (1991): An Unified Data Structure based on linear Quadtree, Department of Photogrammetry and Remote Sensing, WTUSM, Wuhan, 1991.

Liedtke, C.-E.; M. Ender (1989): Wissensbasierte Bildverarbeitung, Springer-Verlag Berlin Heidelberg, 1989.

McKeown, D.M., Jr.(1984): Knowledge-based Aerial Photo Interpretation, Photogrammetria, 39:91-123, 1984.

Mooneyhan, D. W.(1983): The Potential of Expert Systems for Remote Sensing, 17th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, 9-13 May 1983.

Nagao, Makoto and Takashi Matsuyama(1980): A structural Analysis of Complex Aerial Photographs, Plenum Press, New York and London, 1980.

Negoita, C.V. (1985): Expert Systems and Fuzzy Systems, The Benjamin/Cummings Publishing Company, Inc., 1985.

Niemann, H. und H. Bunke (1987): Künstliche Intelligenz in Bild- und Sprachanalyse, B. G. Teubner Stuttgart, 1987.

Pan, He Ping (1990): A Spatial Structure Theory in Machine Vision and its Application to Structural and Textural Analysis of Remotely Sensed Images, dissertation in ITC, 1990.

Pentland, Alex P. (1986): From Pixels to Predicates, Recent Advances in Computational and Robotic Vision, Ablex Publishing Corporation, Norwood, New Jersey, 1986.

Riekert, Wolf-Fritz (1990) : Das RESEDA-Projekt: Ein wissensbasierter Ansatz zur Auswertung von Rasterbilddaten im Rahmen eines Umweltinformationssystems, FAW und Siemens AG, Unternehmensbereich Kommunikations- und Datentechnik, 1990.

Rosenfeld, A. and A. C. Kak (1982): Digital Picture Processing, Second Edition, Vol.2, Academic Press, 1982, pp 57-344.

Towers, Simon and R. Baldock (1988): Application of a Knowledge-Based System to the Interpretation of Ultrasound Images, Pattern Recognition & Automation Section, MRC Clinical & Polulation Cytogenetics Unit, Western General Hospital, Crewe Road, Edinburgh, UK., 1988.

Vernazza, Giannil; S. B. Serpico and S. G. Dellepiane (1987): A Knowledge-Based System for Biomedical Image Processing and Recognition, IEEE Transactions on Circuits and Systems, Vol. CAS-34, No. 11, 11. 1987.

Willkomm, Philipp (1991): Extracting Vector Data and Attributes from Stereo Images with PHOCUS, EARSeL Workshop on Relationship of Remote Sensing and Geographic Information Systems, Hannover, 9. 1991.

Xu, Yonglong (1988): Knowledge Engineering and Photogrammetry Tomorrow, Co. 3, the 16th ISPRS Congress, Kyoto, Japan, July 1988.

Xu, Yonglong (1991a): suchbasierte Verfahren zur automatischen Raster-Vektor und Vektor-Raster Transformation der Grauwertbilder, Institut für Photogrammetrie und Ingenieurvermessung, Universität Hannover, 1.1991.

Xu, Yonglong (1991b): Untersuchung zur Bildsegmentation zwecks der nachfolgenden wissensbasierten Objekt-Erkennung, Institut für Photogrammetrie und Ingenieurvermessung, Universität Hannover, 7.1991.

Zhang, Wenxing (1986) : The Basic Principles and Methodologies of Artificial Intelligence, lecture documents, WTUSM, Wuhan, 1986.

Zhang, Wenxing (1987) : Artificial Intelligence and Mapping Today, lecture documents, WTUSM, Wuhan, 1987.