A System Environment for Knowledge Based Image Analysis

H.Niemann
Lehrstuhl für Informatik 5 (Mustererkennung)
Universität Erlangen Nürnberg
Martensstraße 3
8520 Erlangen, F.R. of Germany

The purpose of *image analysis* is the automatic computation of a *symbolic description* of an image, or an image sequence. The content of the symbolic description must correspond to the requirements of a particular *application* and provide a description on an adequate level of abstraction. In industrial scenes, for example, a description of the quality of manufactured objects or special tasks for a robot is required. To achieve this, complex applications require the explicit representation of extensive task-specific knowledge about objects that have to be recognized, their relations to each other, and the properties of their task domain; this makes up the model. Furthermore there is a need for control algorithms to use the stored knowledge during the analysis process. The control algorithms have to be able to deal with imprecise terms of knowledge like 'dark, big, or long' and such terms must be represented by the formalism. The efficiency of an image analysis system therefore depends on both the knowledge base and the control algorithm. Besides the two modules *knowledge* and *control*, a *data base* for the intermediate results as well as *methods of preprocessing and segmentation* of images are necessary. In addition, it may be useful to have facilities for *knowledge acquisition, explanation*, and a transparent *user interface.*

In most systems processing proceeds from the sensor signal, i.e. the image, represented by an array of integer sample values, via different levels of more and more abstract representations of the content of the image. *Representational levels* may be, for example, edges and regions obtained from an initial segmentation, three-dimensional surfaces, symbolic names for objects and their relations, conceptions of motion or events like 'placing a part into a device for assembly'. Each level of abstraction can be characterized by the knowledge available at that level, by the type of processes which can be activated at the level, and by the assigned structures where a process at some level produces structures of the next level of abstraction at its output. In order to get an efficient and flexible strategy for the understanding process, the intergration of the different levels in one *homogeneous system architecture* is a useful approach.

Our approach to obtain a homogeneous system environment is to use a *semantic network* which is tailored to the requirements of pattern analysis. Semantic networks are special, directed graphs, with labeled nodes and links. But only certain types of nodes and links are allowed. The basic idea is quite simple: Information about conceptions (objects, relations, events, situations, facts) are represented by nodes, interrelationships between nodes are represented by labeled links to also mark the kind of the relationship. Such well structured networks with their capability to also handle procedural knowledge are an adequate way to build up knowledge bases for image (and speech) analysis systems. A complete knowledge based system can be integrated into such a structure. The idea is that the semantic network with its nodes and links is a kind of skeleton specifying and representing the overall structure of the declarative knowledge base. Attached to each node there is local task-specific procedural knowledge. The declarative and procedural knowledge is activated by a problem independent control algorithm.

Three different types of nodes exist in the network. First of all, a *concept* gives the intensional description of an object, an event, or classes of objects or events. The second type, called *instance*, represents a hypothesis that a real world object or event satisfies the definition of a concept. The concepts form the knowledge base, while their extensional sets, the instances, are the intermediate and final results. The third type, called *modified-concept*, is a concept that is restricted because of existing instances of other concepts. Relationships between concepts (and also between instances and between modified concepts) can be built up by three different types of *links*. They are denoted by specialization, part, and concrete. The links define a hierarchy in the network. *Attributes* characterize physical properties of a concept, like height or color, while the *structural relations* show concept dependent relationships between parts, concretes, and attributes. A software environment ERNEST is available for designing a semantic network of this type. A procedural semantics is defined for the ERNEST network language which is independent from the content of a knowledge base. It is built up of six *rules* which only look for concepts, for modified concepts, and for instances in their premise. Given some concepts as potential goals of an analysis process, the recursive application of the rules builds up the skeleton of the *search space* of an analysis process. This skeleton itself is a tree because in some situation more than one rule or one rule with different premises may be activated. Competing intermediate results which are created by the different attached procedures split up this skeleton into the complete search space. As a strategy for the search process the A*-algorithm is used.

The knowledge representation language, the six rules for instantiating concepts, and the control stategy make up the system environment ERNEST for pattern analysis. It is in use in both speech and *image understanding*. As an example of an application we outline its use in analysis of industrial scenes.

The goal of our work in *industrial scene analysis* is the supervision of a production cell for a manufacturing process. This requires recognition and localization of objects, inspection of their quality, and detection of unforeseen events. There are three main subtasks involved: initial image segmentation, knowledge acquisition, and knowledge based analysis of an actual image. We will give some details of the first two subtasks.

Image segmentation deals with the problem of decomposing an image into parts that are relevant to the other components of the system. For example, contours, corners, regions and informations about the three-dimensional structure of surfaces are extracted from the image. The results are represented in segmentation objects, like lines and corners, their attributes and interrelations. We focus on some results based on edge detection, region growing, and shape from shading; the latter point being a first step to incorporating three-dimensional segmentation results. A state of the art algorithm for edge detection, based on concepts of Canny is implemented in the image processing system. This operator optimizes three criteria: detection, localization, and single response to an edge. Of course, the detection of edge points is only the first step to image segmentation. The edge points have to be connected to lines and then detection of corners and vertices, for example, is neccessary.

In contrast to edge detection, where discontinuities in the image are of interest, region operators are searching for homogeneities in the image. This criterion for homogeneity has to be specified. For example, the assumption could be made that there are only small changes of the gray value within an object or an interesting part of it. All connected points of the image where the criterion is met are grouped to a region. Since regions have a closed contour, one has not to deal with the problem of connecting line segments, as is neccessary when using edge detectors. On the other hand it is very difficult to specify a global criterion for homogeneity that is convenient for all parts of the image.

To analyse images of natural scenes it is very important to obtain three-dimensional information from an image, where in particular the curvature of surfaces is of interest. We have implemented different methods of shape from shading to retrieve this information from an image. These methods of shape from shading only use a single gray scale image and analyse the changes of the gray value to calculate the surface normal at each point of the image. To do this, some assumptions have to be made. In general the image intensity depends on four factors. They are the illumination of the scene, the reflectance of the surface, the viewpoint, and the geometry of the surface. The algorithm is quite robust as we obtain a good reconstruction of the surfaces although the surface reflectance is not Lambertian.

Results of segmentation are used by a knowledge based analysis process. The representation of declarative knolwedge is done in the concepts of a semantic network. The ability of automatic *knowledge acquisition (learning)* is very important for modern image analysis systems.

As was already mentioned above there are several levels of abstraction represented in the knowledge base and there is always one level which is close to the input data. Learning the knowledge on this level inductively from input data (examples) is a great facilitation. But on the other hand it is not efficient to learn the knowledge of a high level of abstraction from examples (e.g. learning an action plan of a robot from image sequences of real actions). This type of knowledge should be transferred into the knowledge base deductively using an explicit description (e.g. an action plan created by off line programming). There may be the problem that the high level knowledge is not explicitly formulated but that it does exist in the mind of some experts. In such a case at first this knowledge has to be elicitated.

A general knowledge acquisition algorithm for supervised learning of a class of situations or for learning of single situations will be described. This learning can be done inductively and/or deductively. It uses a semantic network for knowledge representation. The goal of the knowledge acquisition algorithm is an almost automatic construction of part of the declarative knowledge base for an image analysis system. The algorithm is an integral part of the ERNEST environment.

The main idea of the ERNEST built in *knowledge acquisition algorithm* is that learning is done gradually. The first step is observation and description. Any sensor data has to be transformed into an internal representation. The second step is comparison of the formed descriptions. Correspondences are searched for whenever a new observation is made. During the third step the gained information will be used to build up a model for image analysis. An optional fourth step may compare the created model with negative examples. The goal of this step is to verify and to specialize the model. The input data of the algorithm may be positive examples and/or descriptions provided by CAD files. When learning the shape of industrial objects, the images of these objects are the positive examples whereas the CAD data of these objects form a description which can be used too.

The knowledge base of the acquisition process is based on the semantic network language. It is separated into a declarative and a procedural part. The declarative part is a so called *model-scheme* which is formed by concepts. It specifies the structure of the facts that you want to learn. All detectable primitives and relations between those primitives are defined by the model-scheme. In case of learning 3D-objects, the model-scheme defines that a 3D-object consists of surfaces, edges, and vertices. The procedural part of the knowledge of the acquisition process consists of the procedures required to perfom the above mentioned three steps of the acquisition process.

The following references quote some previous and related work as well as additional details concerning the above text.

References

[BIR 88a] T.Birnthaler: Entwurf und Implementierung von Prozeduren für die automatische Wissensakquisition. Studienarbeit, Lehrstuhl für Informatik 5 (Mustererkennung) Universität Erlangen-Nürnberg 1988

[BRA 79a] R.J.Brachman: On the Epistemological Status of Semantic Networks. In [FIN 79a] 3-50

[BRA 85a] R.J.Brachman, I.A.Schmolze: An Overview of the KL-ONE Knowledge Representation Language. Cognitive Science, Vol.9 (1985) 171-216

[BRU 88a] H.Brünig, HNiemann: IPAX: Konzeption und Realisierung eines Bildverarbeitungssystems, Proc. 10. DAGM-Symposium, Informatik Fachberichte 180, Springer Berlin Heidelberg New York 1988, 60-67

[CAN 86a] J.Canny: A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No.6, November 1986, 679-698

[DIE 81a] T.Dietterich, R.Michalsky: Inductive Learning of Structural Descriptions. AI, vol. 16 (1981) 257-294

[EIC 86a] W.Eichhorn, H.Niemann: A Bidirectional Control Strategy in a Hierarchical Knowledge Structure. Proc. 8. Int. Conf. on Pattern Recognition, Paris France 1986, 181-183

[FIN 79a] N.V. Findler (ed.): Associative Networks. Academic Press, New York 1979

[FRA 87a] R.T.Frankot, R. Chellappa: A Method for Enforcing integrability in Shape from Shading Algorithms, Proc. First Int. Conf. on Computer Vision, London, 1987, 118-127

[HOR 74a] S.L.Horowitz, T. Pavlidis: Picture Segmentation by a Directed Split-and-merge Procedure, Proc. 2nd InternatiAnal Joint Conference on Pattern Recognition, Copenhagen, 1974, 424-433

[HOR 77a] B.K.P.Horn: Obtaining Shape from Shading Information, Winston, P., The Psychology of Computer Vision, McGraw-Hill, New-York, 1977, 115-155

[KUM 87a] F.Kummert, H.Niemann, G.Sagerer, S.Schröder: Werkzeuge zur modellgesteuerten Bildanalyse und Wissensakqusition — Das System ERNEST. In M. Paul (Hrsg.): GI - 17. Jahrestagung Computerintegrierter Arbeitsplatz im Büro. Springer Berlin Heidelberg New York Tokyo 1987, 556-570

[LEE 85a] C.-H.Lee, A. Rosenfeld: Improved Methods of Estimating Shape from Shading Using the , Light Source Coordinate System, Artificial Intelligence Vol. 26 (1985) 125-143

[LEV 79a] H.Levesque, J.Mylopoulos: A Procedural Semantics for Semantic Networks. In [FIN 79a] 93-121

[MYL 83a] J.Mylopoulos, H.Levesque: An Overview of Knowledge Representation. In M.Brodie, J.Mylopoulos, J.V.Schmidt (eds): On Conceptual Modelling: Perspectives from Artificial Intelligence, Databases, and Programming Languages. Springer New York Berlin Heidelberg Tokyo 1983

[NIE 81a] H.Niemann: Pattern Analysis. Springer Berlin Heidelberg New York 1981

[NIE 85a] H.Niemann: A Homogeneous Architecture for Knowledge Based Image Understanding. Proc. 2. Conf. Artificial Intelligence Applications. Miami Florida 1985, 88-93

[NIE 87a] H.Niemann, H.Bunke: Künstliche Intelligenz in Bild- und Sprachanalyse. Teubner, Stuttgart 1987

[NIE 87b] H.Niemann: Wissensbasierte Verarbeitung industrieller Szenen. In K.Feldmann, M.Geiger, U.Herzog, H.Niemann, B.Schmidt, H.Wedekind: Proc. Fachtagung „Rechnerintegrierte Produktionssysteme", Universität Erlangen-Nürnberg, Erlangen 1987, 133-147

[NIL 82a] N.J.Nilsson: Principles of Artificial Intelligence. Springer Berlin Heidelberg New York 1982

[PER 86a] W.A.Perkins: INSPECTOR: A Computer Vision System that Learns to Inspect Parts. IEEE Trans. Pattern Aanlysis and Machine Intelligence PAMI-5 (1986) 584-592

[SAG 85a] G.Sagerer: Darstellung und Nutzung von Expertenwissen für ein Bildanalysesystem. Informatik Fachberichte 104, Springer Berlin Heidelberg New York 1985

[SRO 88a] S.Schröder, H.Niemann, G.Sagerer: Knowledge Acquisition for a Knowledge Based Image Analysis System. Proc. of the European Knowledge Acquisition Workshop (EKAW88), GMD-Studien 143, Bonn 1988, 29-1 - 29-15