Lecture Notes in Computer Science 2048

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Josef Pauli

Learning-Based Robot Vision

Principles and Applications

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Author

Josef Pauli
Christian-Albrecht Universiät zu Kiel
Institut für Informatik und Praktische Mathematik
Preusserstr. 1-9, 24105 Kiel, Germany
E-mail: jpa@ks.informatik.uni-kiel.de

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Pauli, Josef:
Learning-based robot vision : principles and applications / Josef Pauli. - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2001
(Lecture notes in computer science ; 2048)
ISBN 3-540-42108-4

CR Subject Classification (1998): I.4, I.2.9-11, I.2.6

ISSN 0302-9743
ISBN 3-540-42108-4 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Typesetting: Camera-ready by author, data conversion by Boller Mediendesign
Printed on acid-free paper SPIN 10781470 06/3142 5 4 3 2 1 0

Preface

Industrial robots carry out simple tasks in customized environments for which it is typical that nearly all effector movements can be planned during an off-line phase. A continual control based on sensory feedback is at most necessary at effector positions near target locations utilizing torque or haptic sensors. It is desirable to develop new-generation robots showing higher degrees of autonomy for solving high-level deliberate tasks in natural and dynamic environments. Obviously, camera-equipped robot systems, which take and process images and make use of the visual data, can solve more sophisticated robotic tasks. The development of a (semi-) autonomous camera-equipped robot must be grounded on an infrastructure, based on which the system can acquire and/or adapt task-relevant competences autonomously. This infrastructure consists of technical equipment to support the presentation of real world training samples, various learning mechanisms for automatically acquiring function approximations, and testing methods for evaluating the quality of the learned functions. Accordingly, to develop autonomous camera-equipped robot systems one must first demonstrate relevant objects, critical situations, and purposive situation-action pairs in an experimental phase prior to the application phase. Secondly, the learning mechanisms are responsible for acquiring image operators and mechanisms of visual feedback control based on supervised experiences in the task-relevant, real environment.

This paradigm of learning-based development leads to the concepts of compatibilities and manifolds. Compatibilities are general constraints on the process of image formation which hold more or less under task-relevant or accidental variations of the imaging conditions. Based on learned degrees of compatibilities, one can choose those image operators together with parametrizations, which are expected to be most adequate for treating the underlying task. On the other hand, significant variations of image features are represented as manifolds. They may originate from changes in the spatial relation among robot effectors, cameras, and environmental objects. Learned manifolds are the basis for acquiring image operators for task-relevant object or situation recognition. The image operators are constituents of task-specific, behavioral modules which integrate deliberate strategies and visual feedback control. The guiding line for system development is that the resulting behaviors should meet requirements such as task-relevance, robustness, flexibility,

time limitation, *etc.* simultaneously. All principles to be presented in the work are based on real scenes of man-made objects and a multi-component robot system consisting of robot arm, head, and vehicle. A high-level application is presented that includes sub-tasks such as localizing, approaching, grasping, and carrying objects.

Acknowledgements

Since 1993 I have been a member of the Kognitive Systeme Gruppe in Kiel. I am most grateful to G. Sommer, the head of this group, for the continual advice and support. I have learned so much from his wide spectrum of scientific experience.

A former version of this book has been submitted and accepted as habilitation thesis at the Institut für Informatik und Praktische Mathematik, Technische Fakultät of the Christian-Albrechts-Universität, in Kiel. I'm very grateful to the six persons who have been responsible for assessing the work. These are: V. Hlavac from the Technical University of Prague in the Czech Republic, R. Klette from the University of Auckland in New Zealand, C.-E. Liedtke from the Universität Hannover, G. Sommer and A. Srivastav from the Christian-Albrechts-Universität Kiel, and F. Wahl from the Technische Universität Braunschweig. Deepest thanks also for the great interest in my work.

I appreciate the discussions with former and present colleagues J. Bruske, T. Bülow, K. Daniilidis, M. Felsberg, M. Hansen, N. Krüger, V. Krüger, U. Mahlmeister, C. Perwass, B. Rosenhahn, W. Yu, and my students M. Benkwitz, A. Bunten, S. Kunze, F. Lempelius, M. Päschke, A. Schmidt, W. Timm, and J. Tröster.

Technical support was provided by A. Bunten, G. Diesner, and H. Schmidt. My thanks to private individuals have been expressed personally.

March 2001 *Josef Pauli*

Contents

1. Introduction

The first chapter presents an extensive introduction to the book by starting with the motivation. Next, the *Robot Vision* paradigm is characterized and confronted with the field of *Computer Vision*. Robot Vision is the indisputable kernel of *Autonomous Camera-Equipped Robot Systems*. For the development of such new-generation robot systems the important role of visual demonstration and learning is explained. The final section gives an overview to the chapters of the book.

1.1 Need for New-Generation Robot Systems

We briefly describe present state and problems of robotics, give an outlook on trends of research and development, and summarize the specific novelty contributed in this book.

Present State of Robotics

Industrial robots carry out recurring simple tasks in a fast, accurate and reliable manner. This is typically the case in applications of series production. The environment is customized in relation to a fixed location and volume occupied by the robot and/or the robot is built such that certain spatial relations with a fixed environment are kept. Task-relevant effector trajectories must be planned perfectly during an offline phase and unexpected events must not occur during the subsequent online phase. Close-range sensors are utilized (if at all) for a careful control of the effectors at the target positions. Generally, sophisticated perception techniques and learning mechanisms, *e.g.* involving Computer Vision and Neural Networks, are unnecessary due to customized relations between robot and environment.

In the nineteen eighties and nineties impressive progress has been achieved in supporting the development and programming of industrial robots.

- CAD (Computer Aided Design) tools are used for convenient and rapid designing of the hardware of robot components, for example, shape and size of manipulator links, degrees-of-freedom of manipulator joints, *etc.*

J. Pauli: Learning-Based Robot Vision, LNCS 2048, pp. 1-24, 2001.

- Application-specific signal processors are responsible for the control of the motors of the joints and thus cope with the dynamics of articulated robots. By solving the inverse kinematics the effectors can be positioned up to sub-millimeter accuracy, and the accuracy does not degrade even for high frequencies of repetition.
- High-level robot programming languages are available to develop specific programs for executing certain effector trajectories. There are several methodologies to automate the work of programming.
- Teach-in techniques rely on a demonstration of an effector trajectory, which is executed using a control panel, and the course of effector coordinates is memorized and transformed into a sequence of program-steps.
- Automatic planning systems are available which generate robot programs for assembly or disassembly tasks, *e.g.* sequences of movement steps of the effector for assembling complex objects out of components. These systems assume that initial state and desired state of the task are known accurately.
- Appropriate control mechanisms are applicable to fine-control the effectors at the target locations. It is based on sensory feedback from close-range sensors, *e.g.* torque or haptic sensors.

This development kit consisting of tools, techniques and mechanisms is widely available for industrial robots. Despite of that, there are serious limitations concerning the possible application areas of industrial robots. In the following, problems and requirements in robotics are summarized, which serves as a motivation for the need for advanced robot systems. The mentioned development kit will be a part of a more extensive infrastructure which is necessary for the creation and application of new-generation robots.

Problems and Requirements in Robotics

The lack of a camera subsystem and of a conception for making extensive use of environmental data is a source of many limitations in industrial robots.

- In the long term an insidious wear of robot components will influence the manufacturing process in an unfavourable manner which may lead to unusable products. For exceptional cases the robot effector may even damage certain environmental components of the manufacturing plant.
- Exceptional, non-deterministic incidents with the robot or in the environment, *e.g.* break of the effector or dislocated object arrangement, need to be recognized automatically in order to stop the robot and/or adapt the planned actions.
- In series production the variance of geometric attributes must be tight respectively from object to object in the succession, *e.g.* nearly constant size, position, and orientation. Applications of frequently changing situations, *e.g.* due to the object variety, can not be treated by industrial robots.
- The mentioned limitations will cause additional costs which contribute to the overall manufacturing expenses. These costs can be traced back to

the production of unusable products, the loss of production due to offline adaptation, the damage of robot equipment, *etc.*

The main methodology to overcome these problems is to perceive the environment continually and make use of the reconstructed spatial relations between robot effector and target objects. In addition to the close-range sensors one substantially needs long-range perception devices such as video, laser, infrared, and ultrasonic cameras. The long-range characteristic of cameras is appreciated for early measuring effector-object relations in order to adapt the effector movement timely (if needed). The specific limitations and constraints, which are inherent in the different perception devices, can be compensated by a fusion of the different image modalities. Furthermore, it is advantageous to utilize steerable cameras which provide the opportunity to control external and internal degrees-of-freedom such as pan, tilt, vergence, apperture, focus, and zoom. Image analysis is the basic means for the primary goal of reconstructing the effector-object relations, but also the prerequisite for the secondary goal of information fusion and camera control. To be really useful, the image analysis system must extract purposive information in the available slice of time.

The application of camera-equipped robots (in contrast to blind industrial robots) could lead to damage prevention, flexibility increase, cost reduction, *etc.* However, the extraction of relevant image information and the construction of adequate image-motor mappings for robot control causes tremendous difficulties. Generally, it is hard if not impossible to proof the correctness of reconstructed scene information and the goal-orientedness of image-motor mappings. This is the reason why the development and application of camera-equipped robots is restricted to (practically oriented) research institutes. So far, industries still avoid their application, apart from some exceptional cases. The components and dynamics of more or less natural environments are too complex and therefore imponderabilities will occur which can not be considered in advance. More concretely, quite often the procedures to be programmed for image analysis and robot control are inadequate, non-stable, inflexible, and inefficient. Consequently, the development and application of new-generation robots must be grounded on a learning paradigm. For supporting the development of autonomous camera-equipped robot systems, the nascent robot system must be embedded in an infrastructure, based on which the system can learn task-relevant image operators and image-motor mappings. In addition to that, the robot system must be willing to make life-long experience and adapt the behaviors for new environments.

New Application Areas for Camera-Equipped Robot Systems

In our opinion, leading-edge robotics institutes agree with the presented catalog of problems and requirements. Pursuing the mission to strive for autonomous robots each institute individually focuses on and treats some of

the problems in detail. As a result, new-generation robot systems are being developed, which can be regarded as exciting prototype solutions. Frequently, the robots show increased robustness and flexibility for tasks which need to be solved in non-customized environments. The robots behaviors are purposive despite of large variations of environmental situations or even in cases of exceptional, non-deterministic incidents. Consequently, by applying new-generation robots to classical tasks (up to now performed by industrial robots), it should be possible to relax the customizing of the environment. For example, the manufacturing process can be organized more flexible with the purpose of increasing the product variety.[1]

Beyond manufacturing plants, which are typical environments of industrial robots, the camera-equipped robots should be able to work purposive in completely different (more natural) environments and carrying out new categories of tasks. Examples of such tasks include supporting disabled persons at home, cleaning rooms in office buildings, doing work in hazardous environments, automatic modeling of real objects or scenes, *etc.* These tasks have in common that objects or scenes must be detected in the images and reconstructed with greater or lesser degree of detail. For this purpose the agility of a camera-equipped robot is exploited in order to take environmental images under controlled camera motion. The advantages are manifold, for example, take a degenerate view to simplify specific inspection tasks, take various images under several poses to support and verify object recognition, take an image sequence under continual view variation for complete object reconstruction.

The previous discussion presented an idea of the wide spectrum of potential application areas for camera-equipped robot systems. Unfortunately, despite of encouraging successes achieved by robotics institutes, there are still tremendous difficulties in creating really usable camera-equipped robot systems. In practical applications these robot systems are lacking correct and goal-oriented image-motor mappings. This finding can be traced back to the lack of correctness of image processing, feature extraction, and reconstructed scene information. We have to have new conceptions for the development and evaluation of image analysis methods and image-motor mappings.

Contribution and Novelty of This Book

This work introduces a practical methodology for developing autonomous camera-equipped robot systems which are intended to solve high-level, deliberate tasks. The development is grounded on an infrastructure, based on

[1] The german engineer newspaper VDI-Nachrichten reported in the issue of February 18, 2000, that BMW intends to make investments of 30 billion Deutsche Marks for the development of highly flexible manufacturing plants. A major aim is to develop and apply more flexible robots which should be able to simultaneously build different versions of BMW cars on each manufacturing plant. The spectrum of car variety must not be limited by unflexible manufacturing plants, but should only depend on specific demands on the market.

which the system can learn competences by interaction with the real task-relevant world. The infrastructure consists of technical equipment to support the demonstration of real world training samples, various learning mechanisms for automatically acquiring function approximations, and testing methods for evaluating the quality of the learned functions. Accordingly, the application phase must be preceded by an experimental phase in order to construct image operators and servoing procedures, on which the task-solving process mainly relies. Visual demonstration and neural learning is the backbone for acquiring the situated competences in the real environment.

This paradigm of *learning-based development* distinguishes between two learnable categories: compatibilities and manifolds. Compatibilities are general constraints on the process of image formation, which do hold to a certain degree. Based on learned degrees of compatibilities, one can choose those image operators together with parametrizations, which are expected to be most adequate for treating the underlying task. On the other hand, significant variations of image features are represented as manifolds. They may originate from changes in the spatial relation among robot effectors, cameras, and environmental objects. Learned manifolds are the basis for acquiring image operators for task-relevant object or situation recognition. The image operators are constituents of task-specific, behavioral modules which integrate deliberate strategies and visual feedback control. As a summary, useful functions for image processing and robot control can be developed on the basis of learned compatibilities and manifolds.

The practicality of this development methodology has been verified in several applications. In the book, we present a structured application that includes high-level sub-tasks such as localizing, approaching, grasping, and carrying objects.

1.2 Paradigms of Computer Vision (CV) and Robot Vision (RV)

The section cites well-known definitions of Computer Vision and characterizes the new methodology of Robot Vision.

1.2.1 Characterization of Computer Vision

Almost 20 years ago, Ballard and Brown introduced a definition for the term *Computer Vision* which was commonly accepted until present time [11].

Definition 1.1 (Computer Vision, according to Ballard) *Computer Vision is the construction of explicit, meaningful descriptions of physical objects from images. Image processing, which studies image-to-image transformations, is the basis for explicit description building. The challenge of Computer Vision is one of explicitness. Explicit descriptions are a prerequisite for recognizing, manipulating, and thinking about objects.*

In the nineteen eighties and early nineties the research on *Artificial Intelligence* influenced the Computer Vision community [177]. According to the principle of Artificial Intelligence, both common sense and application-specific knowledge are represented explicitly, and reasoning mechanisms are applied (*e.g.* based on *predicate calculus*) to obtain a *problem solver* for a specific application area [119]. According to this, explicitness is essential in both Artificial Intelligence and Computer Vision. This coherence inspired Haralick and Shapiro to a definition of Computer Vision which uses typical terms of Artificial Intelligence [73].

Definition 1.2 (Computer Vision, according to Haralick) *Computer Vision is the combination of image processing, pattern recognition, and artificial intelligence technologies which focuses on the computer analysis of one or more images, taken with a singleband/multiband sensor, or taken in time sequence. The analysis recognizes and locates the position and orientation, and provides a sufficiently detailed symbolic description or recognition of those imaged objects deemed to be of interest in the three-dimensional environment. The Computer Vision process often uses geometric modeling and complex knowledge representations in an expectation- or model-based matching or searching methodology. The searching can include bottom-up, top-down, blackboard, hierarchical, and heterarchical control strategies.*

Main Issues of Computer Vision

The latter definition proposes to use Artificial Intelligence technologies for solving problems of representation and reasoning. The interesting objects must be extracted from the image leading to a description of the 2D image situation. Based on that, the 3D world situation must be derived. At least four main issues are left open and have to be treated in any Computer Vision system.

1. Which types of representation for 3D world situations are appropriate ?
2. Where do the models for detection of 2D image situations originate ?
3. Which reasoning or matching techniques are appropriate for detection tasks ?
4. How should the gap between 2D image and 3D world situations be bridged ?

Non-realistic Desires in Computer Vision

This paradigm of Computer Vision resembles the enthusiastic work in the nineteen sixties on developing a *General Problem Solver* [118]. Nowadays, the

efforts for a General Problem Solver appear hopeless and ridiculous, and it is similarly ridiculous to strive for a *General Vision System*, which is supposed to solve any specific vision task [2]. Taking the four main issues of Computer Vision into account, a general system would have to include the following four characteristics.

1. A unifying representation framework for dealing with various representations of signals and symbols.
2. Common modeling tools for acquiring models, *e.g.* for reconstruction from images or for generation of CAD data.
3. General reasoning techniques (*e.g.* in fuzzy logic) for extracting relevant image structures, or general matching procedures for recognizing image structures.
4. General imaging theories to model the mapping from 3D world into 2D images (executed by the cameras).

Continuing with the train of thought, a General Vision System would have to be designed as a shell. This is quite similar to *Expert System Shells* which include general facilities of knowledge representation and reasoning. Various categories of knowledge, ranging from specific scene/task knowledge to general knowledge about the use of image processing libraries, are supposed to be acquired and filled into the shell on demand. Crevier and Lepage present an extensive survey of knowledge-based image understanding systems [43], however, they concede that *"genuine general-purpose image processing shells do not yet exist."* In summary, representation frameworks, modeling tools, reasoning and matching techniques, and imaging theories are not available in the required generality.

Favouring Robot Vision in Opposition to Computer Vision

The statement of this book is that the required generality can never be reached, and that degradations in generality are acceptable in practical systems. However, current Computer Vision systems (in industrial use) only work well for specific scenes under specific imaging conditions. Furthermore, this specificity has also influenced the design process, and, consequently, there is no chance to adapt a classical system to different scenes.

New design principles for more general and flexible systems are necessary in order to overcome to a certain extent the large gap between general desire and specific reality.

These principles can be summarized briefly by *animated attention*, *purposive perception*, *visual demonstration*, *compatible perception*, *biased learning*, and *feedback analysis*. The following discussion will reveal that all principles

are closely connected with each other. The succinct term *Robot Vision* is used for systems which take these principles into account.[2]

1.2.2 Characterization of Robot Vision

Animated Vision by Attention Control

It is assumed that most of the three-dimensional vision-related applications must be treated by analyzing images at different viewing angles and/or distances [12, 1]. Through exploratory controlled camera movement the system gathers information incrementally, *i.e.* the environment serves as external memory from which to read on demand. This paradigm of animated vision also includes mechanisms of selective attention and space-variant sensing [40]. Generally, a two-part strategy is involved consisting of attention control and detailed treatment of the most interesting places [145, 181]. This approach is a compromise for the trade-off between effort of computations and sensing at high resolution.

Purposive Visual Information

Only that information of the environmental world must be extracted from the images which is relevant for the vision task. The modality of that information can be of quantitative or qualitative nature [4]. In various phases of a Robot Vision task presumably different modalities of information are useful, *e.g.* color information for tracking robot fingers, and geometric information for grasping objects. The minimalism principle emphasizes to solve the task by using features as basic as possible [87], *i.e.* avoiding time-consuming, erroneous data abstraction and high-level image representation.

Symbol Grounding by Visual Demonstration

Models, which represent target situations, will only prove useful if they are acquired in the same way, or under the same circumstances, as when the system perceives the scene in real application [75]. It is important to have a close relation between physically grounded task specifications and the appearance of actual situations [116]. Furthermore, it is easier for a person to specify target situations by demonstrating examples instead of describing visual tasks symbolically. Therefore, visual demonstration overcomes the necessity of determining quantitative theories of image formation.

Perception Compatibility (Geometry/Photometry)

In the imaging process, certain compatibilities hold between the (global) geometric shape of the object surface and the (local) gray value structure in the photometric image [108]. However, there is no one-to-one correspondence

[2] The adequacy will become obvious later on.

between surface discontinuities and extracted gray value edges, *e.g.* due to texture, uniform surface color, or lighting conditions. Consequently, qualitative compatibilities must be exploited, which are generally valid for certain classes of regular objects and certain types of camera objectives, in order to bridge the global-to-local gap of representation.

Biased Learning of Signal Transformation

The signal coming from the imaging process must be transformed into 2D or 3D features, whose meaning depends on the task at hand, *e.g.* serving as motor signal for robot control, or serving as symbolic description for a user. This transformation must be learned on the basis of samples, as there is no theory for determining it *a priori*. Each signal is regarded as a point in an extremely high-dimensional space, and only a very small fraction will be considered by the samples of the transformation [120]. Attention control, visual demonstration, and geometry/photometry compatibilities are taken as bias for determining the transformation, which is restricted to a relevant signal sub-space.

Feedback-Based Autonomous Image Analysis

The analysis algorithms used for signal transformation require the setting or adjustment of parameters [101]. A feedback mechanism is needed to reach autonomy instead of adjusting the parameters interactively [180]. A cyclic process of quality assessment, parameter adjustment, and repeated application of the algorithm can serve as backbone of an automated system [126].

For the vast majority of vision-related tasks only Robot Vision systems can provide pragmatic solutions. The possibility of camera control and selective attention should be exploited for resolving ambiguous situations and for completing task-relevant information. The successful execution of the visual task is critically based on autonomous learning from visual demonstration. The online adaptation of visual procedures takes possible deviations between learned and actual aspects into account. Learning and adaptation are biased under general compatibilities between geometry and photometry of image formation, which are assumed to hold for a category of similar tasks and a category of similar camera objectives.

> General representation frameworks, reasoning techniques, and imaging theories are no longer needed, rather, task-related representations, operators, and calibrations are learned and adapted on demand.

The next Section 1.3 will demonstrate that these principles of Robot Vision are in consensus with new approaches to designing autonomous robot systems.

1.3 Robot Systems versus Autonomous Robot Systems

Robots work in environments which are more or less customized to the dimension and the needs of the robot.

1.3.1 Characterization of a Robot System

Definition 1.3 (Robot System) *A robot system is a mechanical device which can be programmed to move in the environment and handle objects or tools. The hardware consists essentially of an actuator system and a computer system. The actuator system is the mobile and/or agile body which consists of the effector component (exterior of the robot body) and the drive component (interior of the robot body). The effectors physically interact with the environment by steering the motors of the drive. Examples for effectors are the wheels of a mobile robot (robot vehicle) or the gripper of a manipulation robot (manipulator, robot arm). The computer system is composed of general and/or special purpose processors, several kinds of storage, etc., together with a power unit. The software consists of an interpreter for transforming high-level language constructs into an executable form and procedures for solving the inverse kinematics and sending steering signals to the drive system.*

Advanced robot systems are under development which will be equipped with a sensor or camera system for perceiving the environmental scene. Based on perception, the sensor or camera system must impart to the robot an impression of the situation wherein it is working, and thus the robot can take appropriate actions for more flexibly solving a task. The usefulness of the human visual system gives rise to develop robots equipped with video cameras. The video cameras of an advanced robot may or may not be a part of the actuator system.

In camera-equipped systems the robots can be used for two alternative purposes leading to a *robot-supported vision system (robot-for-vision tasks)* or to a *vision-supported robot system (vision-for-robot tasks)*. In the first case, a purposive camera control is the primary goal. For the inspection of objects, factories, or processes, the cameras must be agile for taking appropriate images. A separate actuator system, *i.e.* a so-called *robot head*, is responsible for the control of external and/or internal camera parameters. In the second case, cameras are fastened on a stable tripod (*e.g. eye-off-hand system*) or fastened on an actuator system (*e.g. eye-on-hand system*), and the images are a source of information for the primary goal of executing robot tasks autonomously. For example, a manipulator may handle a tool on the basis of images taken by an eye-off-hand or an eye-on-hand system. In both cases, a dynamic relationship between camera and scene is characteristic, *e.g.* inspecting situations with active camera robots, or handling tools with vision-based manipulator robots. For more complicated applications the cameras must be separately agile in addition to the manipulator robot, *i.e.* having a robot of

its own just for the control of the cameras. For those advanced arrangements, the distinction between robot-supported vision system and vision-supported robot system no longer makes sense, as both types are fused.

The most significant issue in current research on advanced robot systems is to develop an *infrastructure*, based on which a robot system can learn and adapt task-relevant competences autonomously. In the early nineteen nineties, Brooks made clear in a series of papers that the development of autonomous robots must be based on completely new principles [26, 27, 28]. Most importantly, autonomous robots can not emerge by simply combining results from research on Artificial Intelligence and Computer Vision. Research in both fields concentrated on reconstructing symbolic models and reasoning about abstract models, which was quite often irrelevant due to unrealistic assumptions. Instead of that, an intelligent system must interface directly to the real world through perception and action. This challenge can be handled by considering four basic characteristics that are tightly connected with each other, *i.e. situatedness*, *corporeality*, *emergence*, and *competence*. Autonomous robots must be designed and organized into task-solving behaviors, taking the four basic characteristics into account.[3]

1.3.2 Characterization of an Autonomous Robot System

Situatedness

The autonomous robot system solves the tasks in the total complexity of concrete situations of the environmental world. The task-solving process is based on situation descriptions, which must be acquired continually using sensors and/or cameras. Proprioceptive and exteroceptive features of a situation description are established, which must be adequate and relevant for solving the specific robot task at hand. Proprioceptive features describe the internal state of the robot, *e.g.* the coordinates of the tool center point, which can be changed by the inherent degrees of freedom. Exteroceptive features describe aspects in the environmental world and, especially, the relationship between robot and environment, *e.g.* the distance between robot hand and target object. The characteristic of a specific robot task is directly correlated with a certain type of situation description. For example, for robotic object grasping the exteroceptive features describe the geometric relation between the shape of the object and the shape of the grasping fingers. However, for robotic object inspection another type of situation description is relevant, *e.g.* the silhouette contour of the object. Based on the appropriate type of situation description, the autonomous robot system must continually interpret and evaluate the concrete situations correctly.

[3] In contrast to Brooks [27], we prefer the term *corporeality* instead of *embodiment* and *competence* instead of *intelligence*, both replacements seem to be more appropriate (see also Sommer [161]).