VISATEC: Vision-based Integrated Systems Adaptive to Task and Environment with Cognitive abilities

Christian Perwass, Josef Pauli, Gerald Sommer

The aim of the VISATEC project is to further the use of computer vision in industrial applications. Research for this project will therefore concentrate on novel methods for feature extraction, object representation, object recognition, and behavior-based system architectures. The goal is to create a system which offers a set of algorithms and design patterns, such that the system can be easily adapted and may even adapt itself to a given task.

1 Introduction

In many applications, ranging from industrial over military to civil, visual sensory input plays an increasingly important role. In fact, there could be many more applications using visual input, if only appropriate computer vision algorithms would exist. Applications where computer vision is already applied very successfully are the sorting of letters by recognition of (hand-) writing, automated inspection of components, assembly tasks in the automobile industry, etc. Usually, tasks where computer vision can be employed successfully, are those where stringent constraints on the environment can be assumed or enforced. Some aspects that play an important role here are the lighting of a scene, reflective properties of objects in the scene, scene background and camera calibration. Even small changes in any one of these may result in a breakdown of a state of the art computer vision algorithm. The main reasons for these weaknesses may be found in the following aspects.

- *Inadequate representation schemes* for two- and higher-dimensional image signals, for image features at several layers of abstraction, and for cognitive information.
- *Unreliable feature extraction* due to varying influences of the environment or the imaging device as for example, reflections, occlusions, lighting variations, quantization effects, etc.
- *Lack of concepts for the fusion of several cues* of different type. Such a fusion should occur between different layers of abstraction, both within and between images.
- *Missing adaptation and learning mechanisms.* Mechanisms of this type are essential for the construction of robust and adaptable algorithms.
- *Uni-directional, non-dynamic algorithms.* There are no self-organizing task- or error-based feedback loops.

The VISATEC project was created with the goal to advance the state of the art in computer vision with respect to the above mentioned problems, in order to obtain active, adaptable robot vision systems. These weaknesses have also been recognized by other researchers in the field and thus are receiving more and more attention. In this paper, the basic concepts on which the VISATEC project is based that may help to overcome the identified problems, will be presented.

Three parties work jointly on the VISATEC project. The cognitive vision group at the Christian-Albrechts-Universität in Kiel, Germany, with Prof. Dr. Gerald Sommer (scientific coordinator), Dr. Christian Perwass (administrative coordinator), Oliver Granert and Martin Krause. The computer vision laboratory at the university of Linköping, Sweden, with Prof. Dr. Gösta Granlund, Dr. Klas Nordberg, Robert Söderberg and Fredrik Viksten. The industry partner is ISRA Vision Systems AG, Germany.

2 Research Concepts

A fundamental design principle that is to be used in this project, is the behaviour-based design [10, 3]. This implies that each aspect of the project can only be developed in view of all other aspects. In particular, a test-bed is needed not only to check the applicability of the system design in a real life environment, but also to train and organize basic behaviors on which the system is based. The practical application chosen as final demonstrator for this project is a skillful robot arm. This does not refer to the type of robot arms that can be found in most industrial applications, where a robot arm simply follows a pre-programmed path. Instead, the robot arm should adapt or be easily adaptable to the task at hand. Furthermore, a high level of robustness is expected, such that it can work in different and variable environments. It would also be desirable for the arm to decompose a complex task into simpler tasks on its own.

The task on which the robot arm will be tested initially, is to grasp objects out of a box. This application was suggested by the industry partner and is of high practical relevance. It is not too difficult to solve this problem if the environment can be controlled appropriately. In this case that would mean to have the objects in question ordered in a particular way within the box, to control the lighting of the objects and to have a priori knowledge about useful feature points. The vision aspect is then only used to achieve a certain tolerance with respect to the actual position of the box and slight variations in the positions of the objects within the box.

However, the task to be solved within the VISATEC project is more general. Since the objects that are to be detected have to be known beforehand, the system needs to learn the corresponding object representations. It should be as simple as possible for a user to train the system appropriately. For example, the user might simply have to present an object to the system, by placing it on a revolving table. The system then learns a representation of the object that is most suitable. The objects in the box are not ordered in any way. They may have any orientation and will also partly occlude each other. The idea then is to have a robot arm with an attached camera, plus a number of stationary cameras. This system is controlled by a perception-action cycle, whereby the current camera views control the movement of the robot arm. This is used to initially bring the attached camera into a position relative to an object, such that the object and its pose can be recognized. Then the robot arm tries to grasp the object while constantly checking and improving on the object's and its own pose estimate. A successfully grasped object is then placed outside the box and the process starts anew.

Grasping out of a box is only one application a skillful robot arm could be used for. Other examples are a garbage sorting system, a tire unloading system or a household robot arm. In the following, the different aspects of this task that need to be treated will be discussed in some more detail.

2.1 Feature Extraction

In order to represent and recognize objects from image data, it is necessary to extract features from the images. An image itself is initially only a collection of pixels each having a particular gray or color value. Semantic information, however, is encoded in the relation of pixels or pixel regions to each other. Therefore, an object has to be represented by a set of features, which contain such information. One can distinguish between local features, as for example, points, lines, corners, curvatures, and global features, as for example, eigenvectors of a principle or independent component analysis or a Fourier transform. It is important that features should be detectable under different lighting conditions and independent of the orientation in which they appear in an image.

Within the VISATEC project two concepts were regarded as most promising for the task of feature extraction: the structure multivector [2] and the scene tensor [6]. The structure multivector is an embedding of local image structure in a Clifford algebra [8] element. It contains information about the local intrinsical dimensionality of an image, which can be used to detect edges independent of their orientation and corners. The scene tensor also combines local image information about corners, edges, curvatures and symmetries in a somewhat different way.

2.2 Object Representation

Image features as discussed in the previous section, present a first abstraction level from the raw image data. The features extracted by the structure multivector and the scene tensor are local and thus only take into account the immediate neighborhood of a pixel. An object will typically consist of a number of such local features. What distinguishes one object from another is then the particular constellation of features. An object representation using a set of given features should be such that an object can be robustly recognized. This implies that an object representation should be rotation and translation invariant. One way to achieve this are multi-level triplet invariants [5]. Another possibility is the representation of an object by its contour using Fourier descriptors [9]. Yet another way is to learn, in some way, the features present in different views of an object.

2.3 Object Recognition

Depending on the object representation chosen, the object recognition will have to differ somewhat. Nevertheless, it seems sensible to always use some kind of neural network. Many different types of neural networks have been developed. However, when it comes to classification, they all have in common that they attempt to transform the input data in such a way that different classes can be separated linearly. In the VISATEC project radial basis function networks [7], dynamic cell structures [1], support vector machines and channel coding followed by an associative network [4], will be considered in particular.

2.4 Situation Analysis

An important aspect of the VISATEC project is the task-adaptive attentive object and situation analysis. That is, it is not only important to recognize an object, but also to understand how it is situated with respect to other objects. For example, in order to grasp an object it is necessary to have some information about the relative position and orientation of the effector of a robot arm with respect to the object [7]. The use of active vision is of particular importance here. This allows the system to constantly reevaluate the relative positions of all objects in a scene, including the robot arm, such that changes in the environment can be taken into account. Furthermore, a moving camera should improve the robustness of object recognition, and allow for attentive behavior.

2.5 System Architecture

The goal of the VISATEC project is not to produce a solution to the particular problem of grasping objects out of a box. This should only serve as a demonstrator. The system itself should be more general. That is, it should be easily adaptable to new tasks and within a task it should automatically adapt to the environment.

In the final system the user should be able to choose from a large bank of basic algorithms and design patterns. Design patterns should serve as template solution methods for particular classes of tasks [7]. A very general design pattern is, for example, the perception-action cycle. A design pattern may then be filled with content using algorithms from the algorithm bank. For example, using the design pattern 'perception-action cycle' a user may attach a number of cameras as input devices to obtain a number of views of the contents of a box. A neural network may then serve as a processing element to detect objects and a robot control as effector to grasp objects. The system should then be able to go through a learning phase and eventually perform the task of grasping objects out of a box.

To implement such a system architecture, a software package, called PACLib, was developed which is based on a similar idea as CORBA but allows for real time applications (see `www.ks.informatik.uni-kiel.de/~paclib` for more details).

3 Outlook

The VISATEC project started in March 2002 and by now initial concepts have been made more precise and developed further. A good amount of software has been written for the system architecture and many algorithms have been implemented. The project members will therefore soon be in a position to test design patterns and newly developed algorithms in a complete system. A first demonstrator showing the feasibility of the project's concepts will probably be available by March 2003. It is planned that the industry partner starts adapting the project's results to real robotic environments by March 2004, such that by the end of the project in March 2005, a first running prototype system is available.

Literatur

[1] J. Bruske and G. Sommer. Dynamic cell structure learns perfectly topology preserving map. *Neural Computation*, 7:845–865, 1995.

[2] M. Felsberg. *Low Level Image Processing with the Structure Multivector*. PhD thesis, University of Kiel, 2001.

[3] G. Granlund. The complexity of vision. *Signal Processing*, 74:101–126, 1999.

[4] G. Granlund. An Associative Perception-Action Structure Using a Localized Space Variant Information Representation. In *Proceedings of Algebraic Frames for the Perception-Action Cycle (AFPAC)*, Kiel, Germany, September 2000.

[5] G. Granlund and A. Moe. Unrestricted recognition of 3-D objects using multi-level triplet invariants. In *Proceedings of the Cognitive Vision Workshop*, Zürich, Switzerland, September 2002.

[6] K. Nordberg. The structure tensor in projective spaces. Report LiTH-ISY-R-2424, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, April 2002.

[7] J. Pauli. *Learning-based Robot Vision*, volume 2048 of *LNCS*. Springer-Verlag, 2001.

[8] Ch. Perwass. *Applications of Geometric Algebra in Computer Vision*. PhD thesis, University of Cambridge, 2000.

[9] B. Rosenhahn, Ch. Perwass, and G. Sommer. Free-form pose estimation by using twist representations. *Algorithmica*, 2002.

[10] G. Sommer. Algebraic aspects of designing beaviour-based systems. In *Algebraic Frames for the Perception-Action Cycle*, volume 1315 of *LNCS*, pages 1–28. Springer-Verlag, 1997.

Acknowledgments

The VISATEC project is supported by the European Commission in the fifth framework program, IST-2001-34220.

Contact

Prof. Dr. Gerald Sommer, Dr. Christian Perwass
Christian-Albrechts-Universität zu Kiel
Institut für Informatik
Olshausenstr.40, 24098 Kiel
gs, chp @ks.informatik.uni-kiel.de
www.visatec.info

Dr. habil. Josef Pauli
Fraunhofer IITB, Erkennungssysteme
Fraunhoferstr. 1, 76131 Karlsruhe
pauli@iitb.fraunhofer.de

Bild

Christian Perwass studied Physics at the University of London, UK, where he received his Master's degree in 1996. He then moved to Cambridge University, UK, where he received his PhD on *Applications of Geometric Algebra in Computer Vision* in 2000. Since November 1999 he works as PostDoc at the Institut of Informatik at the Universität zu Kiel, Germany. He is currently the administrative coordinator of the VISATEC project.

Josef Pauli, see introductory article.

Bild

Gerald Sommer received a diploma degree in physics from the Friedrich-Schiller-Universität Jena, Germany, in 1969, a Ph.D. degree in physics from the same university in 1975, and a habilitation degree in engineering from the Technische Universität Ilmenau, Germany, in 1988. Since 1993 he is leading the research group cognitive systems at the Christian-Albrechts-Universität zu Kiel, Germany. Currently he is also the scientific coordinator of the VISATEC project.