KNOWLEDGE-BASED 3-D VISION SYSTEMS

Heinz Hügli*, Christian Schütz*, Charles Baur+, Emerico Natonek+

Abstract. The goal of our research is to improve 3-D computer vision by adequate use of knowledge, control and sensing. After a presentation of our approach to 3-D vision, we describe a first experimental configuration that successfully recognises common objects like boxes and therefore demonstrates the feasibility of the approach. Then a more advanced configuration for the recognition of arbitrary objects is proposed and successful recognition experiments are shown. Finally, arriving at the practical impact onto applications, we propose a framework to integrate robotics, virtual reality and 3-D vision in order to perform teleoperation.

1. INTRODUCTION

1.1. Objectives

Our research aims at improving computer vision for applications that involve complex 3-D environments. Three elements are expected to contribute to this improvement: robust knowledge, adequate control, combined range and intensity imaging.

Knowledge basically contributes to the vision by reducing significantly the search process and therefore can greatly improve the vision process. As knowledge base, we use an extensive description of the world and of all objects it contains. While this knowledge is not readily available everywhere, it is already available in a number of applications like advanced manufacturing or teleoperation.

In natural or artificial vision, control mechanisms are various and changing. Bottom-up and top-down are two basic commonly found ways to proceed. In our approach, we propose to combine the two basic approaches by providing both a bottom-up recognition based on range images and a top-down mechanism based on intensity images.

Regarding the nature of the sensing, this research considers both intensity and range imaging. This choice is because range images provide a direct measurement of the 3-D geometry of the scene and therefore have the advantage to provide intrinsic shape features. Significant research work is devoted to the processing and interpretation of these range images.

With above mentioned features, we developed a knowledge based 3-D vision configuration based on the extensive description of the world, a hypothesis generation and verification control scheme and sensing with combined range and intensity images.

1.2. Research project

Major milestones of our research are:

- experimental configuration that recognises common objects like boxes and demonstrates the feasibility of the approach
- advanced 3-D vision approach that extends the capability of above system in its ability to recognise objects of arbitrary shapes
- integration of vision in experimental setups

2. APPROACH TO 3-D VISION

The approach to 3-D vision we investigate and propose uses extended knowledge of the scene and applies a hybrid hypothesis generation and verification control scheme that combines range and intensity imaging.

Key elements of our investigations relate to the range image based 3-D vision environment used to generate hypotheses, the virtual world used as a knowledge base and the matching of real and synthetic images which is used to perform verification.

2.1. Knowledge-based system

According to this approach, the main knowledge is in form of a full description of the world where vision is performed. This description is also known as the virtual world. It includes object models that are adequate for the purpose of vision and topological descriptions.

* Institut de Microtechnique, Université de Neuchâtel, Rue Tivoli 28, CH-2003 Neuchâtel,[hugli,schutz]@imt.unine.ch
+ Institut de microtechnique, DMT, EPFL- Ecublens, CH-1015 Lausanne, [baur,natonek]@imt.dmt.epfl.ch

H. Hügli, Ch.Schütz, Ch. Baur, E. Natonek, "Knowledge-based 3-D Vision Systems", Proc. Priority Programme Informatics Research, Information Conference Module 2, Yverdon, Dec 15-16, 1994

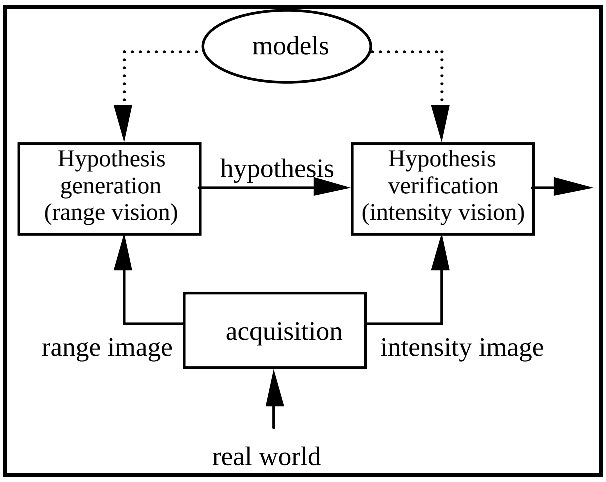


Figure 1: Range and intensity imaging in the hypothesis generation and verification scheme

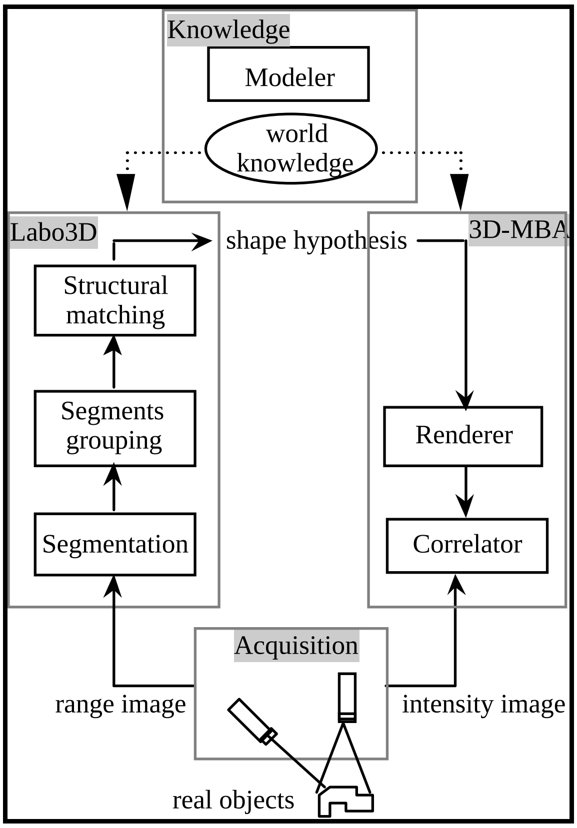


Figure 2: Simplified architecture of the experimental vision system

2.2. Hybrid hypothesis generation and verification

The hybrid hypothesis generation and verification control scheme is illustrated in figure 1. The acquisition module provides information from the real world as both intensity and range images. The range information is used in the hypothesis generation module which is responsible for providing hypotheses, in the form of an estimate of pose and class of an object. The module generates hypotheses in the sense that the solutions it provides are not necessarily unique nor do we expect they are fully correct. This assumption alleviates the constraints on 3-D recognition and permits to make the respective recognition module simple and fast.

Further on the processing path we find the hypothesis verification module. Its purpose is to verify the validity of each emitted hypothesis. It does so by comparing the real intensity image with a synthetic image generated according to the hypothesis interpretation. This interpretation involves the knowledge about objects and world, shortly labelled as models in figure 1. The final result is a set of verified and compatible hypothesises.

2.3. System architecture

Figure 2 shows the overall system architecture, with its four main modules.

The acquisition module delivers range and intensity images. Figure 3 shows typical examples. It comprises a camera for standard imaging and a range imaging device that involves a pair of camera and light projector and operates according to the principle of structured light.

Module Labo3D is a universal development platform with a set of methods and tools performing 3-D object recognition. In the configuration shown, it performs pose estimation from range images and provides the hypothesis about identity and pose of observed objects.

Module 3D-MBA implements a 3-D model-based approach to vision. It provides general means to generate synthetic views of an arbitrary collection of known objects and to compare synthetic and real views. In the configuration shown, it operates in the hypothesis verification mode.

The knowledge module encompasses the object models and the virtual world. Object and world knowledge is in form of a compound model combining vision and rendering parameters; it is implemented as a blackboard.

2.4. Labo3D

Functionally, Labo3D performs 3-D shape recognition from range images [6] [8|. It proceeds in three main steps.

Segmentation. First a segmentation method finds smooth surface patches in the range image by using an algorithm detecting local discontinuities. We call patch each smooth and connected set of rangels (range image element). The patches are always considered flat. Thus a plane is fitted on each patch and segments are built by

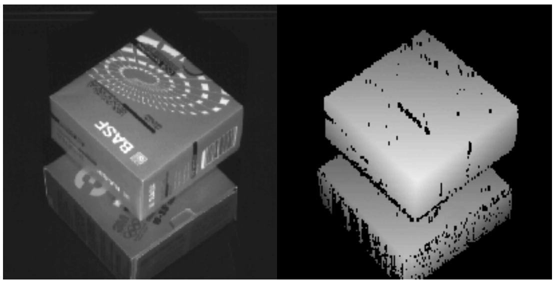
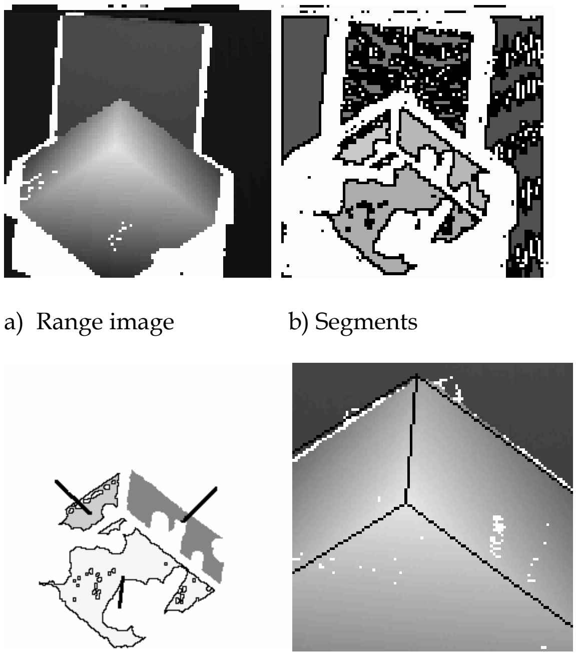


Figure 3: a) Intensity image b) Range image



a) Range image b) Segments

c) Group of segments d) Estimated pose (x 2.5)

Figure 4: Pose estimation from a range image

merging neighboring coplanar patches. This is useful to keep together planar patches that were split into two or several patches in the previous segmentation.

Segment grouping. Then, patches are grouped into sets of few patches to be used as primitives. The idea is to group together neighboring orthogonal segments. The method used looks for orthogonal segments and searches to group them into groups of either 3 or 2 orthogonal segments, which are the primitives.

Structural matching. Finally, a structural matching [1] [2] step searches for correspondence of primitives and models. In the case of box-shaped objects, the method used builds boxes from the primitives and estimates size and pose of the box.

Figure 4 illustrates pose estimation as performed by Labo3D. The example shown provides 5 significant segments from which 3 form a significant group. In d), the estimated box built-up as already described, fits well the original range image.

Robustness to noise is an important topic in these methods, as suggested by this simple example [10].

2.5. Segmentation methods

Segmentation methods for range images were intensively investigated in the frame of this research. The objective is a relative robust method with low complexity. A comparative analysis of methods was performed and led us to selected three different methods for segmenting range images into surface patches. The emphasis is on data-driven segmentation of range images into smooth or planar patches [16] .

One of the selected methods results from the in-depth study of data-driven segmentation methods developed in our lab as a thesis work [10]. The method has the special feature to adapt its parameters to the range image noise and was shown to perform particularly well in a wide class of different range images.

2.6. Modeler

In order to generate object models we need to take into account the most outstanding properties of the real objects. Models are twofold. The first part reflects the geometry of the object. The second part defines what we call the attributes of the object, among which there are the photometric attributes, describing the way the object interacts with light. Further attributes like degrees of freedom and levels of complexity contribute to improve the scene interpretation process.

To generate the geometric database, 3 different techniques have been used. The first uses the conventional technique of defining objects with the keyboard. It is generally considered as tedious. The second and so far the most interesting technique uses 3-D geometric databases of different commercial CAD packages. EXPLORE, the software developed by Thomson Digital Image and used by our system, is able to read different CAD-CAM standards. The advantage of this solution, especially in the case of automated assembly, is that we can take full advantage of the work already done and to rely on a precise and complete database of objects. The third technique relies on the use of 3-D scanners which are used more and more often when the outer surface dimensions of an object have to be determined.

After being modelled, an object is manipulated by a script that allows the user to describe the topology of a scene. The script is the world knowledge database and can be summarise as a set of statements describing the objects, their position in the scene, the position, orientation of camera and lights and so on. We then compose a scene by instancing these objects at various locations. Objects are then manipulated by moving the referential attached to them within a common reference frame.

2.7. 3D-MBA

Rendering. In the synthetic world we can use several different techniques to generate the interaction of light sources with objects. For example, using a method like radiosity, we try to create a perfectly diffuse world. Our system uses more classical render techniques such as scan line algorithms or ray tracing [22].

Correlation. The matching process is the task of finding a set of salient features, in a given image, that matches the model's features. It proceeds in two steps.

Analysing the range image we verify the main feature and we roughly determine the region occupied by the object, its possible dimension and its orientation. By doing so we First we find what we call regions of interest. These are the regions of the image where something is happening. To find them, we verify the main features in the image and we roughly determine the region occupied by the object. These regions will help to speed up the search process by excluding non interesting zones of the image.

Then, the object is rendered and classical correlation is applied between the virtual model and the real scene.

Speed up. For rendering, the geometric database does not always need to be considered entirely. What is necessary is be task dependent. If we know what type of information will be required by the vision system we can set the highest resolution necessary to perform the task. This can be done a priori, on-line or off-line. By a priori we mean that the resolution is fixed during the modelling. We decide that certain given details are not relevant for a search; this means, for example, that either the cost to take them into account or to handle them during the rendering, is too high. Adaptation can also be done on-line either with dedicated techniques which simplify the geometric information or with techniques used during the rendering. The adaptation can also be done off line using the hierarchical representation defined during the modelling. We can then use conventional simplification of the trees encoding the hierarchy.

Several techniques are used to speed up the correlation between virtual and real images. One of the methods, is to choose a salient correlation feature in order to assure a useful recognition based on the correlation results. In figure 5 the logo BASF™ is a salient correlation feature that will ensure the correct object recognition.

3. EXPERIMENTS IN 3-D VISION

3.1. Basic recognition

In this experiment we apply the 3-D vision approach described in figure 2 to recognise boxes in 3-D scenes. The experiment considers floppy boxes of two brands. Figure

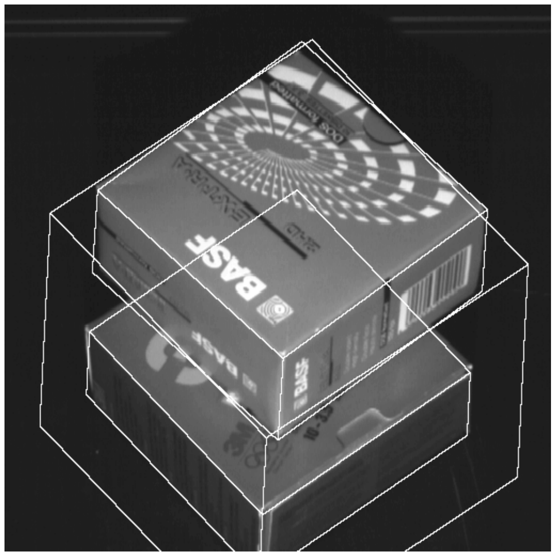


Figure 5: Hypothetical poses

3 is a typical example of a pair of range and intensity images as they result from the acquisition module. We recognise overlapping boxes. Subsequent analysis of the range image gives rise to three hypotheses for boxes as shown in the figure 5. Verification involves correlation tests for all hypotheses and for all the compatible geometrical configurations of two brand of boxes. The two correct hypotheses are accepted and the third is rejected.

3.2. Ambiguous objects

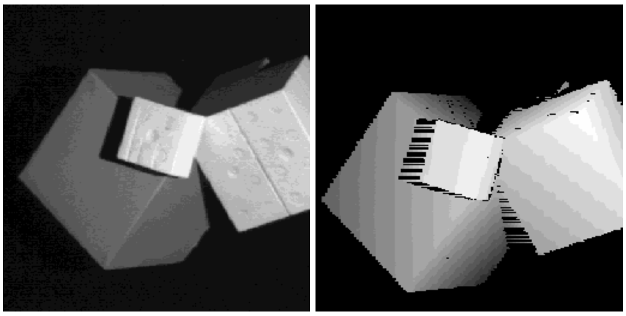
Boxes are among the common objects which are generally ambiguous by their sole shape. Beyond shape, texture information is necessary in order to recognise them. In an experiment to demonstrate the capability of our system to handle ambiguous objects, the system performs standard recognition of a floppy box and displays the corresponding synthetic image, showing the correct position of the printing on the box (figure 6).

3.3. Generic boxes

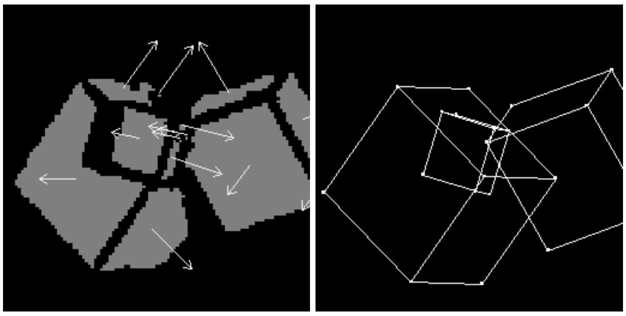
Another experiment demonstrates the recognition of generic boxes, i.e. parallelepipeds of any size. Figure 7 illustrates various steps of the recognition task. Starting from the real scene recorded as range images b) the segmentation process produces the segments represented together with their normal vector in c). Structural matching with the model of a generic box results in the recognised boxes shown in d). This example illustrates the performance of the system to recognise boxes of arbitrary sizes and in arbitrary positions, and more specifically to recognise objects seen under degenerated views, as is the case with the small cube located in the centre of figure 7. The size estimation error of the objects is in the order of 5%.

a) recognised box b) synthetic view

Figure 6: Correct recognition of ambiguous objects



a) intensity image b) range image



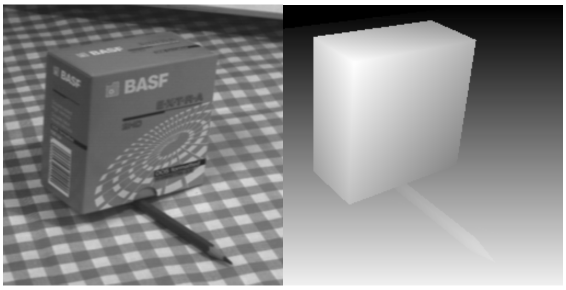
d) segments and normals c) recognised boxes

Figure 7: Recognition of generic boxes

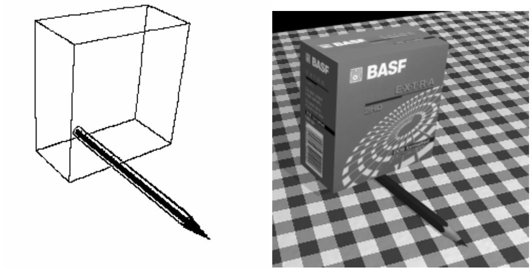
3.4. Recognition of objects on complex backgrounds

Hypothesis verification. To achieve the results shown in figure 8, our system uses prediction-verification processes.

At the range image stage, the system builds a set of hypotheses about the objects present in the scene and then proceeds by trying to confirm/reject them. Newly developed range image segmentation techniques (described above) will extract edges and surfaces for feature checking. If any part of the hypothetical object is missing, the system uses the object model to predict the shape, location and orientation of the missing part. Using the a priori knowledge and the depth map it will extract



a) real image b) range image



c) recognised objects d) synthetic view

Figure 8: Object recognition example using 3D-MBA and range images

the size and exact positioning of both objects (figure 1c). At this step both objects are recognised in a purely geometrical way. For examples see [12]. Next this information is used to construct a more realistic 3-D model (the *virtual image* figure 8d) using the geometrical database and the available photometric attributes. Finally, last recognition is performed by cross correlating the virtual image and the real CCD camera image (figure 8 a).

4. ADVANCED 3-D VISION

4.1. Towards the recognition of arbitrary objects

The recognition approach characterised by the traditional segmentation, grouping and matching scheme presented in figure 2 has a major drawback. Because it relies on the use of robust primitives like the planar patches available in box-shaped objects, and du to the fact that such simple primitives are often simply not available for the objects of interest, the generalisation of this scheme to arbitrary objects is difficult.

An important point of our investigation relates to the possibility of matching arbitrary 3-D shapes. The Iterative closest point algorithm (ICP [20]) being such a tool that allows to compare 3-D shapes directly, without the need to use intermediate representations, we currently study the possibilities to integrate this method into our recognition approach.

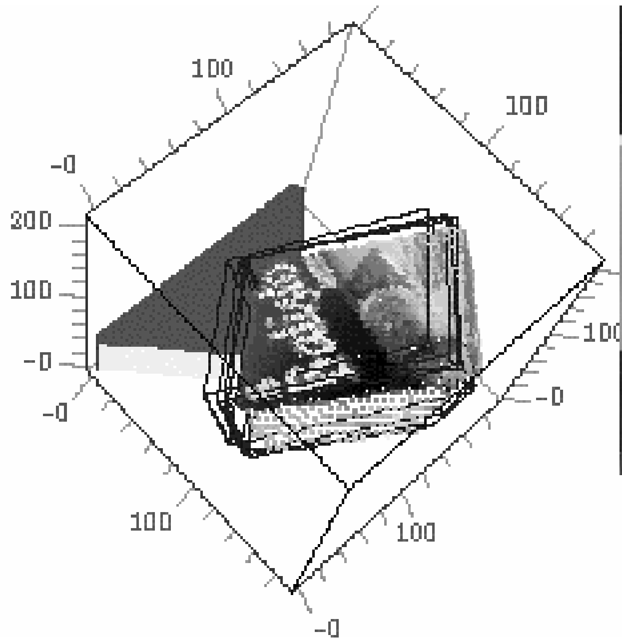


Figure 9: Matching by ICP

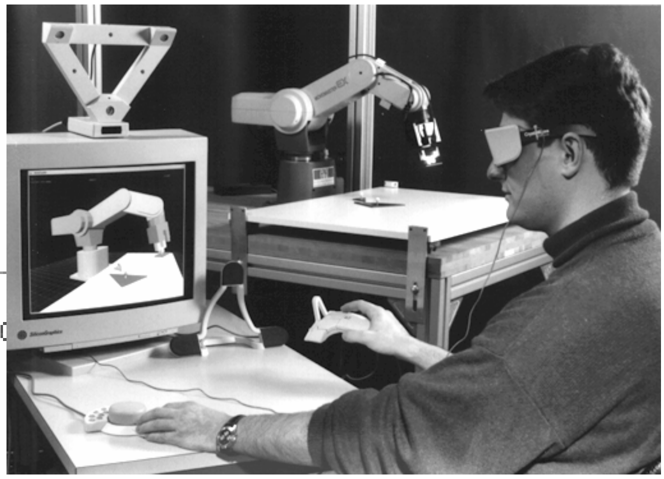


Figure 10: Implemented VR robotic environment (VRRE)

A first study was launched in order to investigate the behaviour and performance of the ICP algorithm in the 2-D recognition tasks. Special attention was given to the study of the recognition of complete and partial shapes. The study considers the practical application of the recognition of puzzle parts [9].

We report experiments were the ICP-algorithm has been used for the recognition of boxes measured by range images. Figure 9 illustrates the result of an experiment where the measured set of point from a range image is matched with a model. In the figure, the Balisto-box represents the experimental data set, and the grey box on the bottom is the model. The wire-frames illustrate the successive positions of the model, as it is iteratively moved onto the measured data set during the execution of the algorithm. This example illustrates a successful recognition and suggests good perspectives for the use of this method for complex objects.

4.2. Towards 3-D vision in VRRE

Reading the phrase of J.C. Craig [23] " In order to make the description of manipulator motion easy for a human user of a robot system, the user shouldn't be required to write down complicated functions of space and time to specify the task " one can easily understand the huge potential of Virtual Reality in a Robotics Environment (VRRE). VRRE gesture via dedicated peripherals can be used for robot guidance or trajectory generation; visual control permits to evaluate the command execution efficiency.

Because a virtual system is only effective if environment changes and object movements are fed-back to the manipulating system, some kind of sensoric feedback is required to perform this task. Virtual reality being a 3-D model of the world, the use of a 3-D model-based vision is highly desirable to help VR to cope with real worlds.

In the field of virtual reality, vision and robotics, we presently have developed an experimental environment that allows to program industrial and mobile robots in a VR environment [19] [13].

It consists of an industrial 5-axis robot, its virtual equivalent and a model-based vision system used in the sensoric feed-back loop. The user is immersed in a 3-D space built out of models of the robot's environment. He directly interacts with the virtual "components", defining tasks and dynamically optimising them. A model based vision system locates objects in the real workspace to update the VRRE through a bi-directional communication link.

In order to enhance the capabilities of the VRRE, a reflex-type behavior based on vision has been implemented. By locally (independently of the VRRE) controlling the real robot, the operator is discharged of small environmental changes due to transmission delays. Thus once the tasks have been optimised on the VRRE, they are sent to the real robot and a semi autonomous process ensures their correct execution thanks to a camera directly mounted on the robot's end effector. On the other hand if the environmental changes are too important, the robot stops, re-actualises the VRRE with the new environmental configuration, and waits for task redesign.

Figure 10 illustrates the developed VRRE. On the left we see the VRRE system together with its 3-D input devices; on the right, we see the real robot, its controller and the model-based vision system. Our vision system has 2 cameras: one on top of the robot's working area, giving a global view of the working space; the other one mounted on the end effector, used for close view analysis. So far the implemented model-based vision is 2-D.

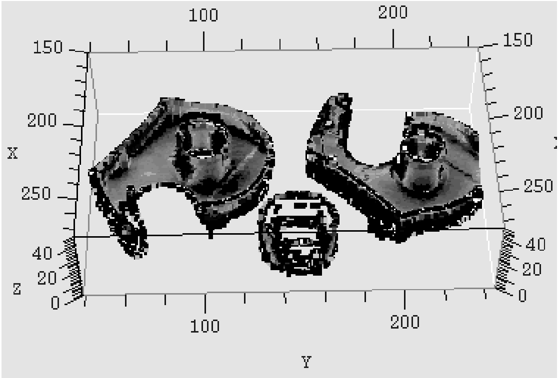


Figure 11: 3-D view of a tape roll

4.3. Towards 3-D vision in assembly

The presented results show the way to improved performance and applicability of knowledge-based 3-D vision. On this way, we arrive next to the problem of recognition and manipulation of 3-D objects in a perspective of application in assembly. Figure 11 represents the range image of a tape roll to be used in next experiments on 3-D vision.

5. VALORISATION

5.1. Publications and presentations

This research was made publicly available by various papers, proceedings, conferences and other means as given below.

Research results were presented at national and international conferences [1] [2] [3] [5] [6] [7] [8] [11] [12] [13]. Many results are also reported in proceedings and in journals as well [1] [2] [3] [6] [7] [8] [11] [12] [13] [14] and two Ph.D thesis are tightly related to the present research work [4] [10]. Additional results are available in technical reports [9] [15] [16] [17] [18] [19] [24].

In addition, results on the exploitation of 3D information have been presented in several laboratories in the USA [15].

5.2. Collaboration with other research teams

The collaboration between the three research teams in Computer vision (T. Pun, CUI-Université de Genève and F. Ade, IKT-ETH Zürich being the two others) is significant and effective. Visits and meetings gathering all participants of the research teams take place at regular intervals. Beyond simple exchange of information, the collaboration concerns the exchange of experience, images, software modules, etc.

Of particular interest is KBVision, a tool for image processing and computer vision that is currently beeing used in the three research teams involved in the Priority Program. Benefits are multiple and various, ranging from software compatibility, ease of communication to software maintenance.

Contacts reagarding this project exist with the NASA Ames Centre (Moffetfield, CA) and Unversité de Clermont Ferrand, France. The common goal is the guidance of mobile and/or conventional robots by 3D vision. A collaboration based on an exchange of researchers is under way.

5.3. Applications

Model-based 3D vision is requested by a large number of applications: interpretation of complex 3-D scenes, surveillance, quality control.

Model-based 3D vision is an ideal complement to virtual reality. The combination of both techniques opens new perspectives for the teleoperation of conventional or mobile robots, and for mechanisms operating on nanoworlds as well [14]. Because the operator interacts with the robotics system at a task oriented high level, VRRE systems are easily portable to other robotics environments (mobile robotics and micro assembly).

5.4. Teleoperation NASA-EPFL

A demonstration of teleoperation organised by IMT-EPFL together with NASA Ames Research Centre (Moffetfield, CA) was held in February 1994. The teleoperation combines vision, robotics and virtual reality. The experiment involves the operation of a real robot from a remote location with tools of virtual reality (VR). At the remote location, the user interacts with the synthetic robot he sees on a screen by means of VR tools.

A particular feature of this demonstration is the use of vision in the vicinity of the robot. Vision is used as a feedback in a system that reports differences between the real and virtual world. The demonstration showed how real objects newly brought in the neighborhood of the robot are recognised and automatically introduced in the virtual world.

Reports about this demonstration were widely diffused and also presented to a wider technical as well as non technical public [5].

6. ACKNOWLEDGEMENT

This research has been funded by the Swiss national Priority Program in Informatics, Knowledge-based systems, under project number 5003-34336.

7. CONCLUSIONS

We proposed and investigated a knowledge-based approach to 3-D vision applying a hybrid control scheme based on range and intensity imaging. The feasibility of this approach is demonstrated by a basic experimental configuration that successfully recognises common objects like boxes. A further step is set with the proposition of an advanced 3-D vision approach that uses generalised matching methods and also with the successful demonstration of its practical use for range images. Finally, the developed VRRE environment constitutes now a testbed for challenging knowledge-based vision tasks.

8. REFERENCES

8.1. Own publications

[1] Amann H.P. & Hügli H. An algorithm for the inexact matching of high level 3D polyhedral representations. In *Selected papers on Model-Based Vision,* editor Hatem Nasr, SPIE Milestone Series, Vol. MS72, 1993

[2] H.P. Amann & H. Hügli. An inexact graph matching algorithm for three-dimensional object recognition. In *Advances in structural and syntactic pattern recognition*, Ed. H. Bunke, Series on Machine Perception and Artificial Intelligence - Vol. 5, pp 421-430, World Scientific Pub. Co., 1993

[3] Baur C. & Natonek E. Model-based vision, robotics and virtual reality. In *Proceeding on VR activity in Europe*, to be published

[4] Baur C. Un système de vision fondé sur les modèles utilisant l'imagerie de synthèse et la corrélation normalisée(MBVS). *Thèse N 998(1992)* Ecole Polytechnique Fédérale de Lausanne.

[5] C. Baur. Gestion d'un système robotisé à l'aide d'une interface basée sur la réalité virtuelle et d'un dispositif de vision artificielle. *Conférence de presse EPFL-NASA*, 16.2.1994

[6] P. Gingins & H. Hügli. Model-based 3-D Object Recognition by a Hybrid Hypothesis Generation and Verification Approach. In *SPIE conf. on Intelligent Robots and Computer Vision*, Boston, Oct. 1994

[7] P. Gingins & F. Vuilleumier. Reconnaissance de formes par des méthodes mixtes: reconnaissance de personages de BD. *Compte-rendus Journée Traitement Image*, Fribourg, Sept. 1993

[8] H. Hügli & P. Gingins. Vision by range and intensity for model-based object recognition. In *2nd Japan-France Congress on Mechatronics*, Takamatsu, Japan, 1-3 Nov. 1994

[10] G. Maître. Segmentation et traitement préliminaire des images de profondeur. *Thèse de doctorat à l' Université de Neuchâtel*, Oct. 1994

[11] E. Natonek and C. Baur. Model based 3-D Object Recognition using Intensity and Range image. In *Proc. SPIE Int. Symposium on Optical Eng. In Aerospace Sensing,* vol 2234-01, Orlando, 1994

[12] Natonek E. & Baur C. Recognition of 3D objects on complex background using model based vision and ranges images. *IEEE Southwest Symp. on Image Analysis and Interpretation.* Dallas, Apr. 21-22, 1944

[13] E. Natonek, L. Flückiger, Th. Zimmerman, C. Baur. Virtual Reality: an intuitive approach to Robotics. In Proc. *SPIE Int. Symposium on Photonics, Telemanipulator and Telepresence Technologies,* vol 2351-32, Boston, 1994

[14] A. Sulzmann, C. Baur. Eine Virtual Reality Schnitstelle für die On / Off-Line *Telemanipulation von Mikrostrukturen.* In Technische Rundschau , April 1994

8.2. Own reports and presentations

[9] H. Hügli & P. Gingins. Expériences de reconnaissance avec l'algorithme ICP. *Rapport IMT 369 HU 09/94*, Université de Neuchâtel, septembre 1994

[15] Baur C. Rapport de voyage US'93. *Rapport interne IMT-DMT*, EPFL, août 1993.

[16] Maître G. & Gingins P. Range image segmentation: performance of three methods based on differential properties. *Report Université de Neuchâtel. IMT 342 HU 06/93,* June 1993

[17] Ph. Gingins. Localisation et mesure de boîtes à partir d'images de profondeur. *Rapport IMT 357 HU 12/93, Université de Neuchâtel*, déc. 1993

[18] L. Claud. Tâches de segmentation sous KBVision. *Rapport IMT HU 94/300, Université de Neuchâtel,* Février 1994

[19] L. Fluckiger. Robotique, Réalité Virtuelle et Vision. *Technical Report IMT-94-01,* EPFL, 1994

[24] C. Schütz & Ph. Gingins. Exploration de la convergence de l'algorithme ICP. *Rapport IMT HU 10/94, Université de Neuchâtel,* octobre 1994

8.3. Other publications

[20] Paul Besl & Neil D. McKay. A Method for Registration of 3-D Shapes. *I.E.E.E. Transactions On Pattern Analysis and Machine Intelligence,* Vol. 14, No 2, pp 239 ff, February 1992.

[21] P. Suetens, P. Fua and A.J. Hanson. Computational Strategies for Object Recognition.

In *ACM Computing Survey*, vol 24 n°1, 1992

[22] D. Thalmann and N. Magnenat-Thalmann. Image Synthesis: Theory and Practice. *Springer-Verlag*, 1987

[23] J.C. Craig. Introduction to Robotics, Mechanics and Control. *Addison Wesley*, 1986