3D Vision in a Virtual Reality Robotics Environment

Christian Schutz*, Emerico Natonek**, Charles Baur**, Heinz Hugli*

*) University of Neuchatel
Institute of Microtechnology
Rue Breguet 2
CH-2000 Neuchatel
Switzerland
Phone: +41 32 718 3455
Fax: +41 32 718 3402
Email: schutz@imt.unine.ch; hugli@imt.unine.ch

**) Swiss Federal Institute of Technology
DMT / IMT
EPFL-Ecublens
CH-1015 Lausanne
Switzerland

ABSTRACT

Virtual reality robotics (VRR) needs sensing feedback from the real environment. To show how advanced 3D vision provides new perspectives to fulfill these needs, this paper presents an architecture and system that integrates hybrid 3D vision and VRR and reports about experiments and results. The first section discusses the advantages of virtual reality in robotics, the potential of a 3D vision system in VRR and the contribution of a knowledge database, robust control and the combination of intensity and range imaging to build such a system. Section two presents the different modules of a hybrid 3D vision architecture based on hypothesis generation and verification. Section three addresses the problem of the recognition of complex, free-form 3D objects and shows how and why the newer approaches based on geometric matching solve the problem. This free-form matching can be efficiently integrated in a VRR system as a hypothesis generation knowledge-based 3D vision system. In the fourth part, we introduce the hypothesis verification based on intensity images which checks object pose and texture. Finally, we show how this system has been implemented and operates in a practical VRR environment used for an assembly task.

KEYWORDS: Virtual reality, virtual reality robotics, 3D vision, knowledge-based vision, range imaging, image correlation

1. INTRODUCTION

Virtual reality robotics (VRR) consists in coupling a virtual reality environment with a robot in order to program its tasks. The programming of robot tasks is often a tedious task which can be greatly improved in user friendliness by using virtual reality (VR) interfaces. The description of the robot motion is done interactively in a virtual space representing the real world setup rather than writing down complicated functions [1, 8]. This allows an off-line programming and verification of the robot tasks.

A perquisite for the success of this approach is a faithful correspondence of virtual and real world. It means that changes in the real world must be fed back into the virtual environment by means of sensing devices. A wide range of sensing devices and methods are candidates for solving application specific tasks. Among them, the recognition of complex 3D objects by vision is still considered as a difficult task, but recent developments in range imaging and 3D interpretation show new perspectives for improvements [6]. If a 3D vision system continuously updates the VR world or part of it, the operator doesn't need to do it laboriously from a camera image of the real world.

Three elements are expected to contribute to the feedback establishment by 3D vision: robust knowledge, adequate control, combined range and intensity imaging. We consider these points as follows during the design of a hybrid 3D vision system:

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 3D geometry of the scene and therefore have the advantage to provide intrinsic shape features.

With above mentioned features, we designed and developed a hybrid 3D vision architecture based on the extensive description of the world, a hypothesis generation and verification control scheme and sensing by range and intensity images. The global architecture is defined in the next section. Following sections describe the implementation of the different architecture modules. Finally, the last section presents the integration of hybrid 3D vision in a VRR environment and illustrates its operation in the frame of an VRR controlled assembly task.

2. HYBRID 3D VISION ARCHITECTURE

2.1. Principle

The architecture for 3D vision we propose uses extended knowledge of the scene and applies a hybrid hypothesis generation and verification control scheme that combines range and intensity imaging in order to obtain reliable object recognition. The models stored in a knowledge database are first used to generate a hypothesis using the range image and second to verify this hypothesis by intensity vision. The architecture and the data flow are illustrated in Fig. 1.

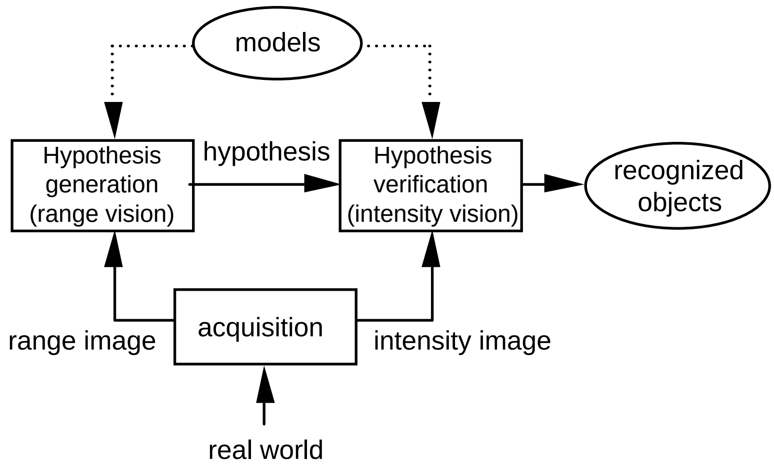

Fig. 1. Range and intensity imaging in the hypothesis generation and verification scheme

The acquisition module provides information from the real world as both intensity and range images. The range information is used for hypothesis generation whereas intensity is used for verification. The object type and pose identification of complex 3D objects is a difficult task when disposing only of one intensity image sensed by a standard CCD camera. The information present in an intensity image is not sufficient to fulfill the recognition task of free-form objects we aim at [3]. Therefore, we choose a system which directly measures the 3D geometry of the scene as obtained with range finders which use for example structured light and active triangulation methods to capture the accurate 3D information of a scene.

The hypothesis generation module provides hypotheses, in the form of an estimate of pose and class of an object. The solutions the module provides are not necessarily unique nor do we expect they are fully correct. This assumption alleviates the constraints on 3D 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 labeled as models in Fig. 1. 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 rendering.

The final result is a set of verified and compatible hypotheses: the recognized objects.

2.2. Modules

We now present in Fig. 2 the overall system architecture in more details, with its four main modules.

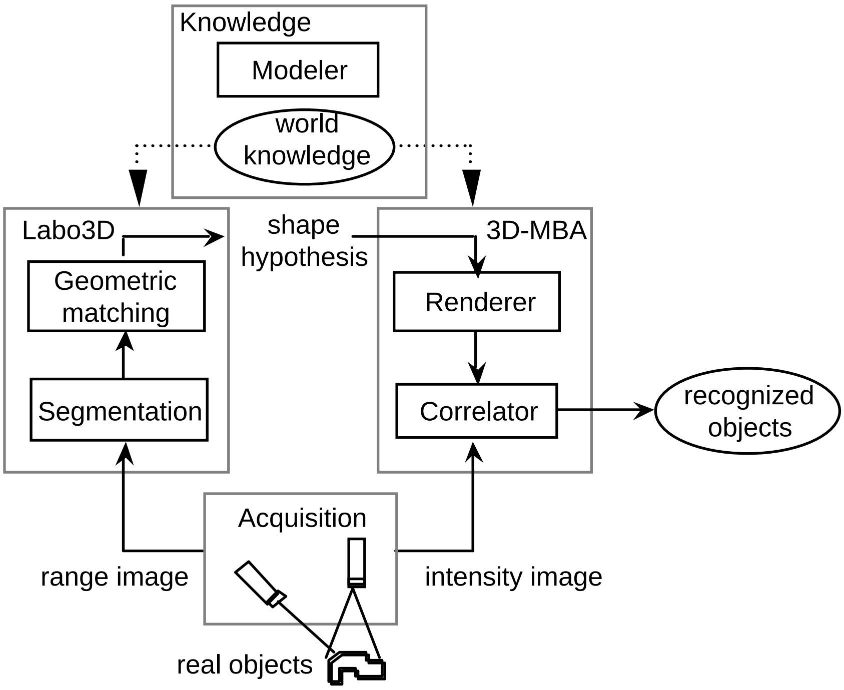

Fig. 2. Different modules of the vision system architecture

The module Labo3D is an universal development platform with a set of methods and tools performing 3D object recognition. In the configuration shown, it performs scene segmentation and pose estimation from range images and provides the hypothesis about identity and pose of observed objects. The different methods are implemented in a graphical programming interface called KBVision developed by Amerinex Applied Imaging.

Module 3D-MBA implements a 3D model-based approach to vision. It provides general means to generate virtual views of an arbitrary collection of known objects and to correlate virtual and real views. In the configuration shown, it operates in the hypothesis verification mode. The application EXPLORE, a software developed by Thomson Digital Image, renders the synthetic images whereas the correlator is implemented on a vision board from Matrox.

The acquisition module senses range and intensity images. The chosen range finder is called BIRIS and produced by VITANA. It consists of a laser stripe and a calibrated standard CCD camera and measures the 3D information of the scene along an one-dimensional profile. The range information is expressed as the camera to object distance and is presented in an one dimensional array. To obtain a complete range image the sensor is moved along an axis perpendicular to the profile. The successive measurements form a 2D range image which allows a direct access to the 3D geometry of the scene. Fig. 3 shows a range image of three tape dispenser parts acquired with the BIRIS range finder.

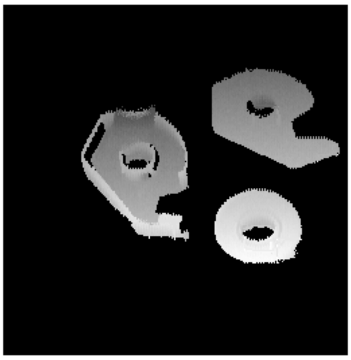

Fig. 3. Range image

The knowledge module encompasses the object models and the virtual world. Object and world knowledge are in form of a compound model combining vision and rendering parameters; it is implemented as a blackboard. Both the Labo3D and the 3D-MBA use the same database to ensure a consistent hypothesis representation. This simplifies the hypothesis generation since both systems use the same reference system. Actually, the two systems reside in different sites and are communicating via the Internet.

2.3. Object models

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, three 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 3D geometric databases of different commercial CAD packages. 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 3D scanners which are used more and more often when the outer surface dimensions of an object have to be determined.

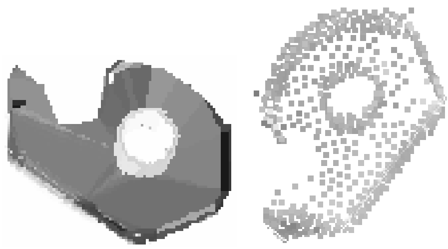

Fig. 4. Different model representations

After being modeled, an object is manipulated by a script that describes the topology of a scene. The script combines informations of the world knowledge database and the generated hypotheses. It can be summarized as a set of statements describing the objects, their position in the scene, the position, orientation of camera and lights and so on.

3. HYPOTHESIS GENERATION

The hypothesis generation module receives a range image of the scene at its input and generates a set of hypotheses which describe the objects present in the scene at its output. It is divided in a segmentation module and a geometric matching module which do object detection and recognition respectively.

3.1. Scene segmentation

Before the different objects are entered into the recognition module they have to be detected and separated from the background. The detection of objects placed on a complex background is possible having several range images showing the scene before and after the object addition or removal. Such objects are located at the regions of significant change in the difference image calculated by the subtraction of two range images from successive acquisitions. As range images may contain non-measured pixels caused by shadows the difference calculation needs a special definition for such cases. Rules to solve this problem are discussed in [10]. The following Fig. 5 shows two typical scenes from an assembly task and the results of the range image difference segmentation showing the added object with a white border.

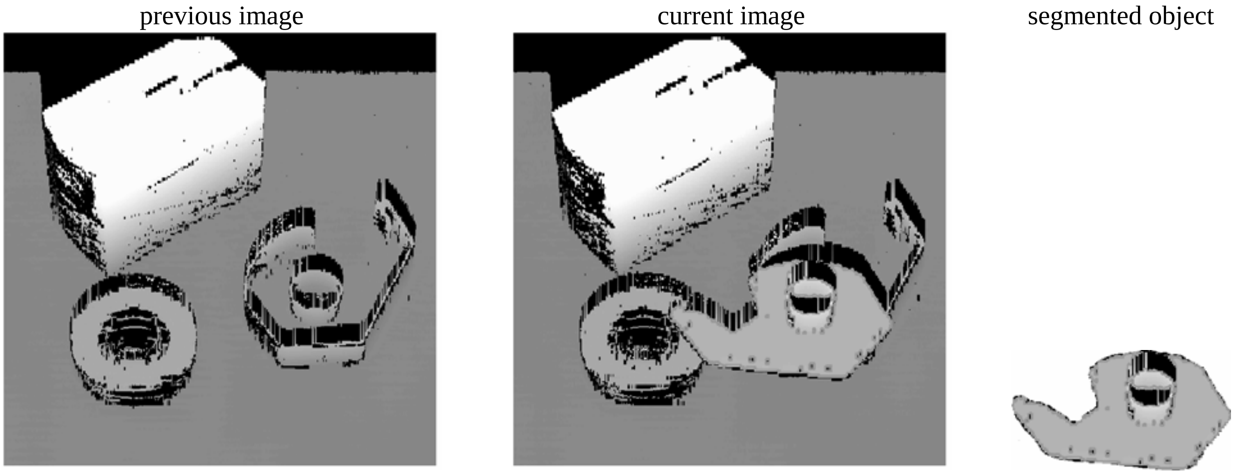

Fig. 5. Segmentation of range image by change detection

3.2. Geometric matching

Because the traditional segmentation and primitive extraction approach works only with very simple objects, we opted here for a recognition principle based on geometric matching. This approach works directly on the 3D coordinates of the object surfaces as obtained from range data. Hence, the recognition is independent on assumptions related to object features. It can be applied to objects of any arbitrary shape: free-form objects.

According to this geometric approach for object recognition, the comparison of test and model surfaces is performed with an iterative closest point matching algorithm (ICP) [2]. In our framework, the test surface is represented by a range image whereas the model surface can be represented by a set of 3D points, polygons or other geometric primitives. For this reason, data obtained from object samples or from CAD models as well can be easily used as models. Consider for the moment that objects are described as set of points in the space. One object is the test; the other is a model. The algorithm proceeds iteratively. First, it pairs every point of the test set with the closest point of the model set. These pairs of closest points between two objects to be matched are then used to calculate the rigid transformation (translation and rotation) which minimizes some distance measurement. The test object is then translated and rotated by the resulting transformation. This procedure is repeated until the predefined distance falls below a threshold τ or the number of iterations exceeds a chosen constant N.

The algorithm converges after some iterations to a solution which is characterized by its distance or error value. During the successive iterations, the test object undergoes successive rigid geometric transformations, bringing it progressively towards a better matching position.

Considering the space of possible initial configurations of the relative test and model pose, successful matching is obtained only for a limited range of it. Since an exhaustive search for successful matching in the space of all possible initial configurations is too expensive, a priori knowledge of the range finder setup is used to reduce the dimension of the search space [9]. The selection of an appropriate number of initial configurations in order to ensure at least one successful matching is discussed in [5].

The test object is being compared to every model. The matching result with lowest errors indicate the recognized object type and since the matching is performed in 3D space the ICP algorithm provides also the relative pose of the test object. Fig. 6a shows a range image of three parts of a tape dispenser. Fig 6b shows the related intensity image with the superimposed point models of the recognized objects in the computed pose.

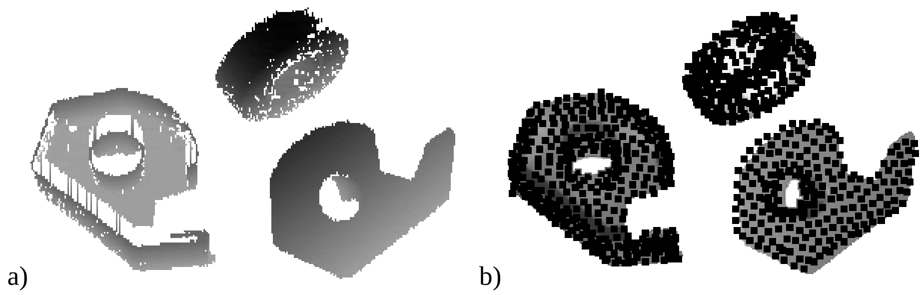

Fig. 6. Range image and recognized models superimposed to the intensity image

At a first glance the necessity of multiple initial configurations seems to introduce much overhead. But since the convergence range of the ICP is relatively large, the number of initial configurations can be kept low [5]. Furthermore, the algorithm converges quickly and allows an estimation of the quality of an initial configuration after few iterations of ICP.

The closest point computation is the most costly part of the ICP algorithm. A search tree method called kD-tree helps to decrease significantly this computation cost. A cost estimation of the complete recognition process using the ICP algorithm and a database with several models can be found in [4].

The possibility to use different representations for models as for example cloud of points, triangulated surfaces or polygons is one of the advantages of the ICP algorithm. Actually, the methods differ only in the way the test to model distance is computed. The remaining parts of the ICP algorithm remain untouched.

A hypothesis contains the type and pose information for an object in the scene. The type is indicated by the corresponding model name and the pose is defined by the translation vector and the rotation matrix transforming the model from the reference point to the object pose in the scene. Since the hypothesis generation and verification module use the same model database the model type name defines an unique model. Furthermore, the pose transformation is well defined since it is applied to a reference point which is attached to each model and used in the hypothesis generation and verification module working in the same world coordinate system.

4. HYPOTHESIS VERIFICATION

The hypothesis verification module uses the generated hypotheses containing the object poses and some a priori knowledge about the scene as camera position, background and robot setup to generate a script describing the complete scene and used for the rendering of the synthetic image. This synthetic image is then compared to the real intensity image in order to verify the validity of the hypotheses.

4.1. Rendering

In the virtual 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 rendering techniques such as scan line algorithms or ray tracing [11]. It transforms the script describing the objects characterized by their geometry and photometry, the light sources and camera view into a synthetic image.

For rendering, the geometric database does not always need to be considered entirely [7]. 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 modeling. We can decide that certain given details are not relevant for a verification; this means, for example, that either the computational 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 modeling. We can then use conventional simplification of the trees encoding the hierarchy.

4.2. Correlation

The correlator compares the synthetic image with the real intensity image in order to verify the correctness of the hypotheses. The actual correlation is performed only in a small correlation site where a model feature is expected to appear. The correlation sites are defined separately for every model during its construction process. The user selects some pertinent regions on a model to be used for the correlation as for example the brand label or corners in the case of a floppy disk box. These regions will help to speed up the search process and improve the reliability by excluding non interesting zones of the image. This local correlation also avoids confusion between objects of the same type present in the scene.

The applied normalized correlation calculates a correlation coefficient for every correlation site. The advantage of the correlation coefficient is its independence to illuminance, a usefull property since it is difficult to model the correct lightening conditions for the rendering.

We performed several experiments with three tape dispenser parts having each two correlation sites attached to it. The use of two sites per object ensures high reliability on pose detection. The sites are chosen on object corners or holes. Fig. 7 shows the matched model features and the corresponding correlation coefficients for three tape dispenser parts.

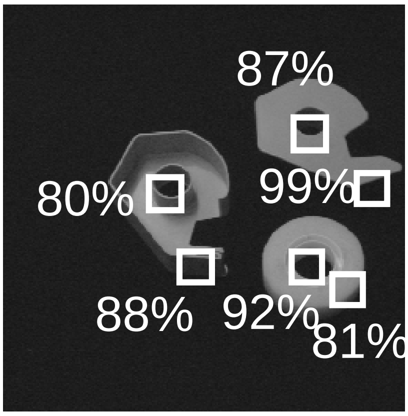

Fig. 7. Correlation coefficients for several regions of interest

Successfully recognized objects have a high correlation coefficient for all their correlation sites. Badly recognized objects may be ignored or sent back to the hypothesis generation module for further inspection. The output of the hybrid 3D vision system is the verified script.

5. 3D VISION INTEGRATION IN VRR

We present the integration and use of the hybrid 3D vision system in a VRR environment. The hybrid 3D vision system sends the verified script to the virtual reality environment which updates its virtual world. This ensures that an up-to-date scene is always presented to the user.

5.1. VRR environment

The virtual world is designed using the tool WTK from Sense8. Dedicated peripherals such as a space mouse are used for robot guidance or trajectory generation. Visual control with stereo glasses permits to evaluate and program robot commands efficiently. Fig. 8 illustrates the VRR environment. On the right we see the virtual reality environment system together with its 3D input devices; on the left, we see the real robot and the working platform. The robot is a Mitsubishi Movemaster RV-M1 with five degrees of freedom.

Fig. 8. VRR environment

The working platform and the robot are directly mapped into the virtual world since their position is known from the workplace setup or the robot position sensors. The hybrid 3D vision updates the virtual scene with the added or removed objects on the working platform.

Once an object appears in the virtual world a grasping trajectory definition becomes easy: the operator only has to point the position to be reached with the space mouse, and to click in order to insert a control point at this position. The robot's orientation is automatically computed. Several tools are also programmed like pick, move, screw, which help the operator.

All these commands are performed off-line which avoids robot damages caused by wrong manipulations. The programmed tasks can be visualized and verified in the virtual world. Finally, the commands are sent to the real robot and executed. After the task completion the vision system updates the virtual world again.

5.2. Application

We choose the task of assembling three tape dispenser parts as shown before in Fig. 3 to demonstrate the operation of the presented hybrid 3D vision system in the above VRR environment. The hybrid 3D vision system detects and recognizes the parts lying on the working platform. Their pose is restricted such that the robot can grasp them but they may touch or cover each another. After hypothesis verification, the recognized parts appear in the virtual world and allow the user to program an assembly task. In our case, the robot has to grasp first the base part of the dispenser and to place it on a assembly template. Then the tape roller is inserted into the base part and finally the cover closes the tape dispenser. A special gripper helps to grip the objects at their holes. Fig. 9 shows two views of the assembly task programed in the VRR environment. The real robot is only used for the final task execution which saves energy and robot resources.

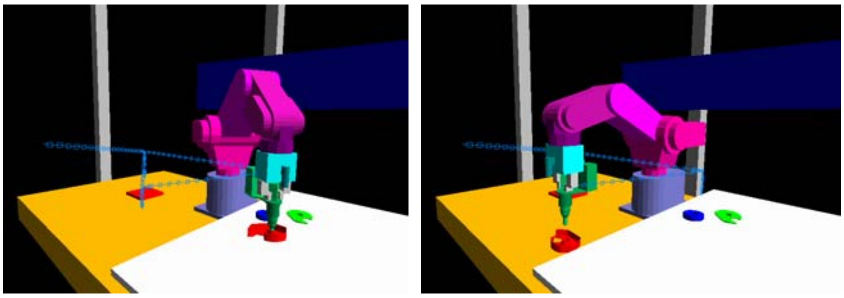

Fig. 9. Robot task programming in the virtual world

Once the programmed task has passed verification by simulation in the virtual world it is downloaded via a serial link to the real robot. The robot executes now the task as shown in Fig. 10.

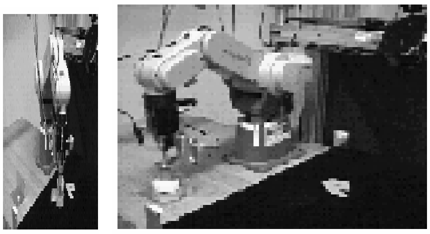

Fig. 10. Robot executing the task programmed in the VRR environment

The system has been tested on several scenes and the results show the successful integration of the hybrid 3D vision system in a VRR environment.

6. CONCLUSIONS

We considered 3D vision as a means to perform sensory feedback in a VRR environment and proposed to apply advanced 3D vision methods based on range imaging and geometric matching as a tool which performs the recognition of arbitrary objects.

A vision system was developed on the basis of a hybrid architecture characterized by a knowledge base and a hypothesis generation and verification structure, two elements that facilitate its integration with the VRR environment. The implemented system was successfully tested in the frame of an assembly task of tape dispenser parts, where vision helps to update the virtual world by recognition and localization of the parts in the real environment.

Experiments with the system showed the very appealing advantages of the applied methods: to treat free-form objects and to integrate well to VRR environments. They speak for the potential and interest of 3D vision in VRR.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

1. J. C. Craig, "Introduction to Robotics," Mechanics and Control, Addison Wesley, 1986.
2. P. Besl and N. McKay, "A method for registration of 3-D shapes," IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239-256, 1992.
3. M. Hebert, J. Ponce, T.E. Boult, A. Gross and D. Forsyth, "Representation for computer vision," Report of the NSF/ARPA workshop on 3D object, CUNY Graduate School and University Center, 1994.
4. H. Hugli, Ch. Schutz, D. Semitekos, "Geometric matching for free-form 3D object recognition", ACCV, Singapore, Vol 3, pp.819-823, 1995.
5. H. Hugli, Ch. Schutz, "How well performs free-form 3d object recognition from range images?", Intelligent Robots and Computer Vision XV, Algorithms, Techniques, Active Vision and Materials Handling, SPIE, Boston, Vol. 2904, 1996.
6. R. C. Jain and A. K. Jain, "Analysis and interpretation of Range Images", Springer Verlag, N.Y., 1990.
7. E. Natonek, "3D-MBA: Système de reconnaissance d'objets par vision basé modèles 3D appliqué aux mondes virtuels robotisés," Thesis No. 1477, Dept. of Microtechnics, Swiss Federal Institute of Technology Lausanne, Switzerland, 1996.
8. E. Natonek, L. Flückiger, Th. Zimmerman, C. Baur, "Virtual Reality: an intuitive approach to Robotics," Proc. SPIE Int. Symposium on Photonics, Telemanipulator and Telepresence Technologies, vol 2351-32, Boston, 1994.
9. Ch. Schutz, H. Hugli, "Towards the recognition of 3D free-form objects," Intelligent Robots and Computer Vision XIV, Algorithms, Techniques, Active Vision and Materials Handling, SPIE, Philadelphia, Vol. 2588, pp. 476-484, 1995.
10. Ch. Schutz and H. Hugli, "Change detection in range imaging for 3D scene segmentation," Europto, Besancon, SPIE 2786, pp. 98-106, 1996.
11. D. Thalmann and N. Magnenat-Thalmann, "Image Synthesis: Theory and Practice," Springer-Verlag, 1987.