Vision-based Control of Assistive Robot FRIEND: Practical Experiences and Design Conclusions

Bildbasierte Regelung des Assistenzroboters FRIEND: Praktische Erfahrungen und Entwurfsentscheidungen

Axel Gräser, Olena Kuzmicheva, Danijela Ristić-Durrant, Saravana K. Natarajan, Christos Fragkopoulos, University of Bremen

Summary In this publication, we describe the evolution of control structures for vision-based control over several generations of the assistive robot FRIEND[1]. We also publish some evaluation results which were the basis of design decisions for the following system generation.

►►► **Zusammenfassung** In diesem Artikel wird die Evolution der bildbasierten Regelung des Assistenzroboters FRIEND über mehrere Generationen beschrieben. Dabei werden Evaluationsergebnisse dargestellt, die die Grundlage für Entwurfsentscheidungen der jeweils nächsten Generation waren.

Keywords Rehabilitation robots, vision-based control, object reconstruction, path planning ►►►
Schlagwörter Rehabilitationsrobotik, Bildbasierte Regelung, Objektrekonstruktion, Trajektorienplanung

1 Introduction

Vision-based control of robots is a key field in the research and application of autonomous robots. There are many different approaches that can be distinguished within this field, including mono/stereo based, position/image based and pure/hybrid visual control [1; 2]. In the control community especially, visual servoing is of particular interest, because it exploits the simple structure of a feedback control loop to provide reduction of system sensitivity to lack of knowledge or inaccuracies in the coordinate transformations between the robot, environment and cameras or imprecision of the robot motion. However, there are real-world robot applications such as object manipulation in cluttered environments (as for the robot FRIEND) in which visual servoing as proposed by Hager has also shortcomings because the robot arm needs to execute specific paths for object manipulation while avoiding obstacles.

In this paper we present the key technologies used and key concepts and algorithms for vision-based control developed over the different generations of the assistive robot FRIEND which belongs to the group of intelligent wheelchair-mounted manipulators and is intended to support disabled people with impairments of their upper limbs in Activities of Daily Living (ADL) and in professional life [3]. We discuss our design decisions which were not only influenced by experiences that we made with specific visual control methods but also by the availability of new technology and by the goal that

[1] FRIEND (Functional Robot arm with user-frIENdly interface for Disabled people)

the robot should become available within a reasonable time frame to its potential users. A key requirement that we focus on in the following discussion is that a useful assistive robot has to carry out complex sequences of actions and not only single isolated actions. We discuss vision-based control structures ranging from classical image-based visual servoing to modified vision-based control structure which allows the introduction of a 7DOF manipulator collision free path planning which is necessary in complex robot scenarios. Since the vision-based control relies on a reliable 3D reconstruction of the robot's environment, we particularly focus on the design of the necessary robust 3D real-world object reconstruction method. That had to be developed in parallel with the development of vision-based control in order to achieve the high dependability required of the complete robot system FRIEND. These robust 3D reconstruction method is based on the introduction of feedback structures at the image segmentation level and on novel 3D modelling of real-world objects. The feedback structures provides robustness of image processing against variable illumination while the novel 3D object modelling provides robustness against unstructured and cluttered scenes as it does not require a priori knowledge on object geometry and it is independent of object texture.

2 Evolution of Vision-Based Control in FRIEND

Figure 1 shows the different generations of the research platform FRIEND that has been developed at the IAT during the last 14 years. In this paper, we concentrate on the different vision-based control methods which are described in the following.

The robot arm of the first prototype FRIEND I was a 6 degrees of freedom (DOF) MANUS robot arm. Since manual control of the robot arm via joystick or by speech control was very tiresome for the user, autonomous positioning of the robot arm was designed. However, due to its lightweight and soft design that generation of the MANUS robot arm had limited accuracy, low repeatability, large backlash and position dependent friction. The position sensors were integrated into the robot base and so were not able to detect position errors of the gripper. In order to deal with these drawbacks of the robot arm, image-based visual servoing was used to control the robot arm.

FRIEND II and the current system FRIEND are equipped with 7 DOF light-weight robot arms which were developed and designed jointly by IAT and Schunk GmbH & Co. KG for use with FRIEND. High reliability, accuracy, repeatability and precise measurements enable the use of the robot measurements directly for control purposes. The improvement of the robot and the need to master manipulation in complex real world scenarios lead to the extension of research from pure visual robot control to advanced trajectory and path planning. In parallel, the camera technology as well as the algorithms for stereo vision, camera calibration and image processing have evolved. That again had significant influence on the overall system design and on the choice of algorithms. Our experiences with FRIEND demonstrated that in everyday support scenarios image processing is the weak point within the robot control structure and that it has to be significantly improved. The idea of feedback structures was introduced in image processing [4] in order to enhance robustness of the vision algorithms and provide reliable and stable information for vision-based reaching and grasping. Considerable resources within the FRIEND program were therefore shifted from visual control to R&D activities in image processing.

2.1 Support Scenarios to be Mastered by FRIEND

For paralysed people who depend on 24 h support by nursing staff, any independence is welcome. In order to justify the high investment in a robot support system, at least 90 min of independence from personal support has to be achieved [5]. The justification of the investment would be strengthened if the disabled person was able to return to professional life with the use of FRIEND and the chance to get the necessary financial support would

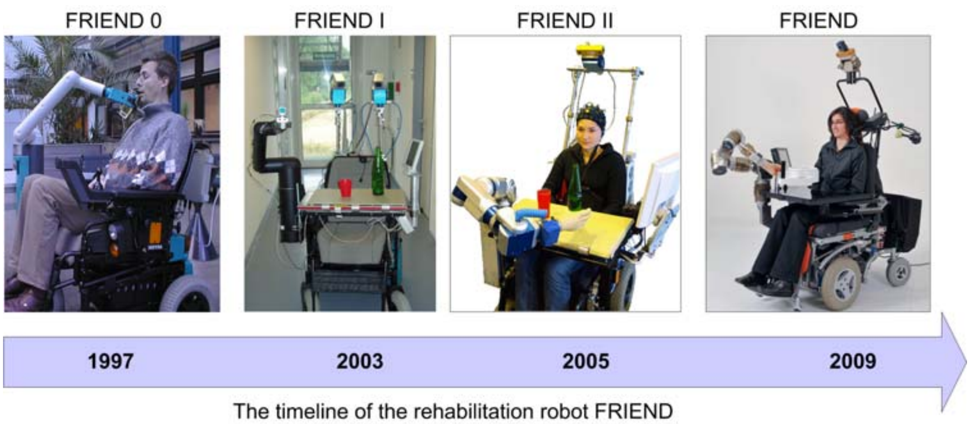

Figure 1 Different generations of research platform FRIEND.


a)

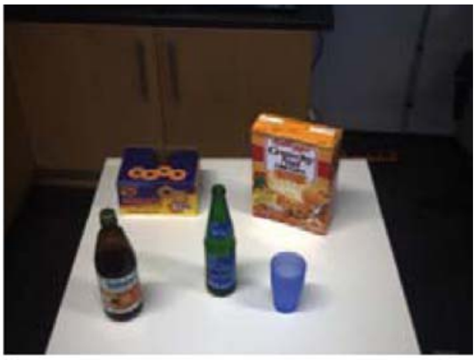
b)

Figure 2 Real-world scenes from FRIEND assistive robot environment in "serving a drink" scenario. (**a**) Refrigerator scene (**b**) Table scene.

increase tremendously. To realize 90 min or more independence from nursing staff for the user it is necessary that FRIEND does not only carry out isolated manipulations but is able to manage a large variety of complete action.

In the different stages of R&D for FRIEND, different support scenarios have been considered. Starting from a simple "serve a drink" scenario in FRIEND I, more advanced scenarios requiring advanced manipulative skills were developed with the next generations of the prototype, FRIEND II, including extended "serving a drink" and "serving a meal" and maintenance scenarios. At the time of writing of this paper, FRIEND is being prepared to support a quadriplegic person for her work in a library where all manual book handling at the specific work place will be carried out by FRIEND.[2]

In order to have a consistent description and also because of limited space, we consider in this paper the evaluation of FRIEND visual perceptual and control capabilities only for the "serving a drink" scenario.

The very first "serving a drink" scenario, realized with FRIEND I, was defined as a fixed sequence of manipulations. The precondition for the scenario is that a bottle and a glass are placed anywhere on the tray and are visible for the camera system. After the glass and bottle are placed on the tray, the user can start the complete scenario with the single speech command "serve a drink". The system executes the task as a sequence of the following actions: grasp the bottle, bring it to a suitable position relative to the glass, pour a drink into the glass, put bottle back on the tray, grasp the glass and move it close to the user's mouth. In this scenario, two essential requirements to provide effective image based visual servoing for the control of the robot arm are: firstly, to reliably recognize the objects (the glass, bottle and gripper) in both 2D camera images during the whole scenario execution, and, secondly, to keep the objects continuously in the field of view of both cameras.

[2] FRIEND is purchased by Schunk as research platform for institutes with an open source software license provided by IAT. FRIEND is also used by IAT in specific research projects with disabled people to verify its benefit in professional and daily life for the users.

In the current generation of FRIEND, this "serving a drink" scenario is embedded in a series of actions which also includes the preparation of a meal in a microwave oven and feeding the user. The drink scenario has now two phases. The first task of the FRIEND's manipulator is to fetch a bottle from a fridge, and to pour a drink from the bottle into a glass placed anywhere on a table. The manipulator then has to put down the bottle, has to fetch the glass and bring it close to the user's mouth.

To allow the manipulator to perform these tasks, it is crucial for the robotic system to perceive its environment visually. In particular, the robot vision system must recognize obstacles and the bottle among other objects inside the fridge and must be robust against changes in illumination. It is necessary to localize the bottle in the fridge in 3D with accuracy high enough to be able to grasp it with the robot. In the second phase of the "serve a drink" scenario, the FRIEND vision system must be able to reconstruct the location of the glass in 3D. For this purpose, it is first necessary to reliably distinguish the glass from the other objects on the table and to correctly recognize it in 2D camera images to enable reliable 3D object reconstruction.

Figure 2 shows images of two scenes from the "serving a drink" scenario. The first imaged scene contains a fridge with various real-world objects placed inside it. Figure 2b shows the scene from the second phase of the "serving a drink" scenario, containing the table with a glass and a bottle placed on it. For the images shown here, both scenes were imaged in the same artificial light conditions. However, the additional light inside the fridge and different background light reflection conditions being present while carrying out the tasks, caused completely different illumination conditions. That results in different appearances of the same bottle in every trial. The vision system therefore has to be robust against different environmental influences. In a cluttered scene, it cannot be assumed that the same objects will always be identically arranged in the target object's neighbourhood. A high robustness against variable lighting conditions is also necessary because lighting and shadow can change over a wide range even while an action is carried out.

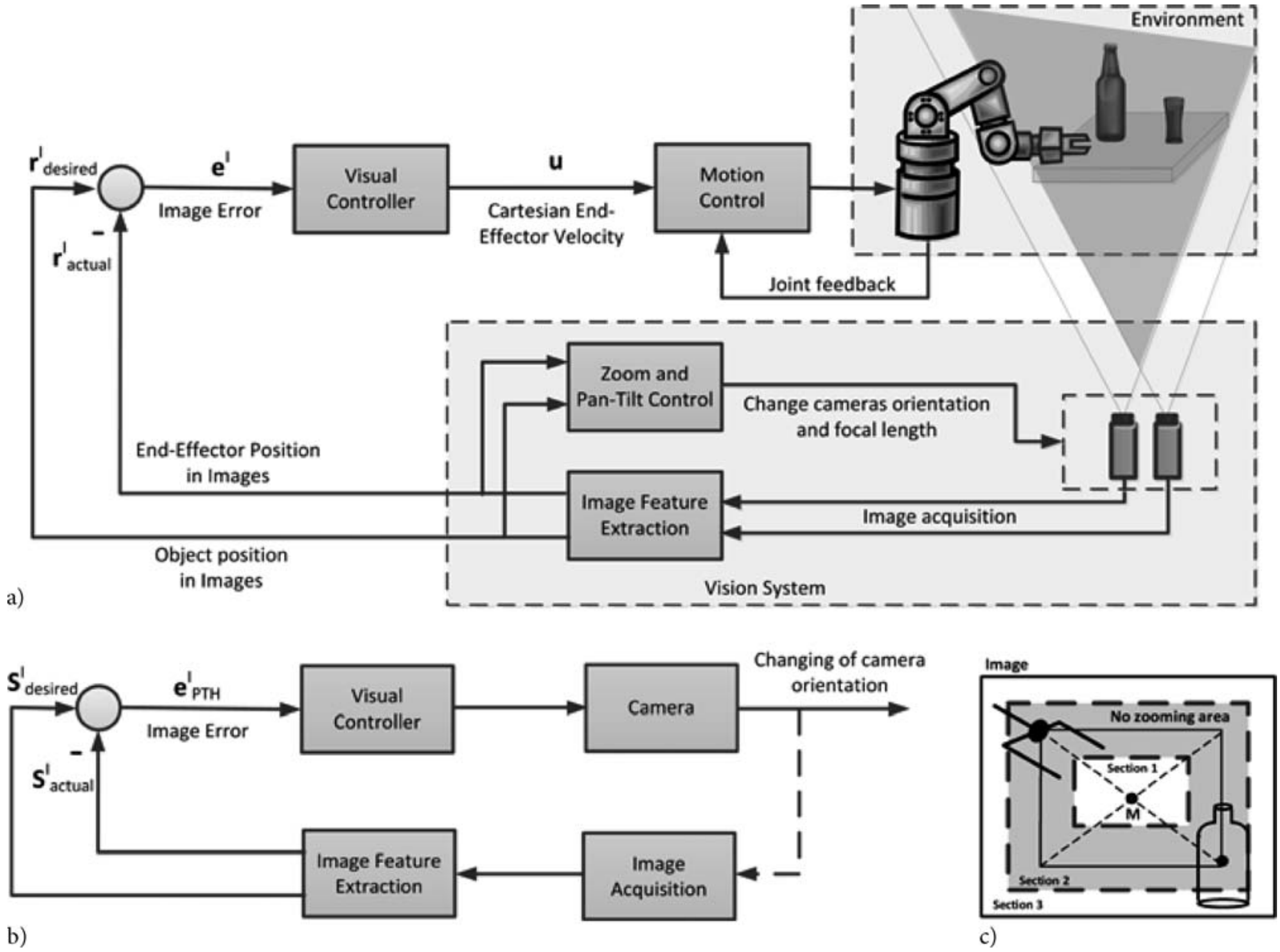

Figure 3 Visual control in FRIEND I. (**a**) Overall control structure. (**b**) Control loop for adjustment of single camera orientation. (**c**) Rules for zoom adjustment.

2.2 Vision-Based Control of MANUS Robot Arm in FRIEND I

The main idea of image-based visual servoing, which was chosen to realize object manipulation within the FRIEND I scenario, is to use the location of objects in the image plane directly as feedback for robot control, see Fig. 3a. The vision system continuously tracks the objects of interest, the glass, the bottle and the robot gripper. The image I control error $\mathbf{e}^I$ is defined as the image distance between the reference point $\mathbf{r}^I_{actual}$ and the target $\mathbf{r}^I_{desired}$. Driving this error to zero in both images captured by stereo cameras is equivalent to driving the reference point, i. e. robot gripper, to the target point, e. g. bottle, in 3D. The visual controller includes the image Jacobian matrix $\mathbf{J}$, which describes the relationship between the Cartesian robot motion and the motion of the robot's image in the camera image frame. A proportional controller is used to calculate the control signal $\mathbf{u}$, i. e. the Cartesian (world W) end-effector velocity $\dot{\mathbf{r}}^W$:

$$\mathbf{u} = \dot{\mathbf{r}}^W = K_p \mathbf{J}^{-1}\left(\mathbf{r}^W, \alpha_{L,R}, \beta_{L,R}, f_{L,R}\right) \mathbf{e}^I \qquad (1)$$

where K_p – is a constant, $\mathbf{J}^{-1}$ – (pseudo)inversed image Jacobian matrix, $\alpha_{L,R}$, $\beta_{L,R}$ – pan and tilt angles of left (L) and right (R) camera accordingly, $f_{L,R}$ – focal length of the left/right camera.

For the system described, successful control is only possible if both the reference point (robot's end-effector) and the target (object to be manipulated) are visible and recognized in both camera images. To satisfy the first condition in a wide range of possible locations, the control of camera parameters, $\alpha_{L,R}$, $\beta_{L,R}$ (pan and tilt angles and $f_{L,R}$ (focal length) had to be included. After the initial object recognition was successful and the objects (bottle, glass) were found, an image-based control loop could be applied to adjust both cameras orientation and focal length. The principle structure of the control loop for orientation adjustment is presented in Fig. 3b.

The control error $\mathbf{e}^I_{PTH}$ is defined for each camera as the difference between the current position of the tracking image feature $\mathbf{S}^I_{actual} = [u,\ v]^T$ and the coordinates of the image centre $\mathbf{S}^I_{desired} = [u_0,\ v_0]^T$:

$$\mathbf{e}^I_{PTH} = \left[u - u_0 \quad v - v_0\right]^T \qquad (2)$$

The control signal is calculated using a proportional controller and an inversed image Jacobian matrix for the pan-tilt heads $\mathbf{J}^{-1}_{PTH}$ which describes the relationship between pixel motions in images $[\dot{u}_L, \dot{v}_L, \dot{u}_R, \dot{v}_R]^T$ and changes in camera orientations $[\dot{\alpha}_L, \dot{\beta}_L, \dot{\alpha}_R, \dot{\beta}_R]^T$, as in [7]:

$$\mathbf{u}_{PTH} = \begin{bmatrix} \dot{\alpha}_L \\ \dot{\beta}_L \\ \dot{\alpha}_R \\ \dot{\beta}_R \end{bmatrix} = K_{PTH}\mathbf{J}_{PTH}^{-1} \begin{bmatrix} \dot{u}_L \\ \dot{v}_L \\ \dot{u}_R \\ \dot{v}_R \end{bmatrix} \quad (3)$$

The adjustment of the focal length enables the scaling of the objects in the image so that the imaged size of the object may be kept almost constant during the control sequence, simplifying the object identification. Moreover, due to the control of the focal length, the adaptation of the field of view of the camera allows the objects of interest to be kept visible by the cameras. To avoid continuously zooming in and out, the zoom control loop includes a rule based controller. Simple rules are used for zooming, see Fig. 3c:

- If object and gripper are in section 1 (centre) then increase zoom
- If object and gripper are in section 2 (grey) then freeze zoom
- If object and gripper are in section 3 (boundary) then decrease zoom.

For successful task realization through visual servoing, reliable object recognition is of major significance. In addition, reliable visual servoing in the system FRIEND requires a particular frame rate, because of the hardware requirements of the controller of the MANUS robot arm. The sampling time of the MANUS controller is 60 ms. Since WINDOWS® is no real-time operating system, an additional response time of about 40 ms must be added if cycle time is considered. Therefore, maximal 10 fps are possible which limits the control results considerably. As described below, additional measurement is then necessary to support the visual control loop.

To facilitate the process of object detection and localization, during the early stage of development of the FRIEND vision system, artificial markers were used. In order to avoid these artificial object markers in FRIEND I, colour based object detection, using colours as natural characteristics of the object that clearly distinguishes it from other objects, was employed [6]. At that stage, the objects used within the scenario, glass, bottle and gripper marker, had different but consistent colours. To provide the required high frame rate despite the restrictions of the used computing hardware (CPU with 2800 MHz), the objects in the image are approximated (see Fig. 4) as ellipses which are described by:

- Image coordinates of the ellipse centre
- Major and minor axis
- Orientation of the ellipse.

The data which are required for visual servoing are dependent on the task to be performed. In case of "grasp a bottle" task, the reference point is defined as the centre of the gripper which is marked by an LED and the target is defined as a virtual point above the large ellipse [6] which describes the bottle.

As described above, the visual perception capability of FRIEND I system was quite restrictive and could not provide all information required for a purely vision based execution of the complete scenario. Hence a smart tray was developed and integrated into the system to support the robot vision [7]. The smart tray provides weight and position information on the objects which are placed on it. These measurements have a much smaller delay than the visual measurements and are used to control the pour in process and to recognize the objects and the free space on the tray. The weight changes which are measured by the scale are used to detect the contact between bottle and tray during the "put down" action immediately, to stop the robot movement and to open the gripper. The inclusion of the tray measurements enhanced the reliability of the "serve a drink" process considerably.

To evaluate the results of scenario execution and to identify remaining problems the "serve a drink" scenario was carried out for 70 trials (40 mornings and 30 afternoons for different illumination). For each trial the bottle and the glass were placed on arbitrary positions on the tray and the task execution was started. The user could hold, cancel or restart the execution at any time. Table 1 summarizes the results.

It could be concluded from the results that the scenario was successfully executed in multiple trials despite weak calibration of the system and that the low positioning accuracy of the robot arm was well compensated through image based visual servoing. However, the results underlined the strong dependence of vision based control on reliable object recognition. If the object recognition was incorrect or incomplete, the task execution failed as well. Object recognition was limited to the recognition of

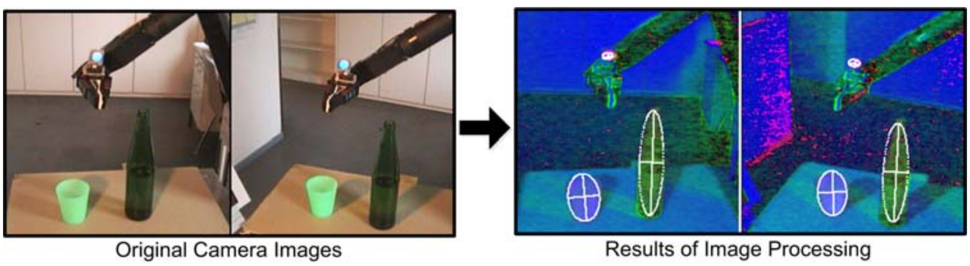

Figure 4 Object description in FRIEND I.

Table 1 Results of reliability tests [8].

	mornings	afternoons	total
Total number of trials	40	30	70
Successful trials	33	20	53
	6**	6**	12**
Totally fail	1	4	5
Object identification fails	1	2	3
Hardware failure	–	2	2

** due to failure the object recognition task was cancelled but successfully executed after intervention by user

small cylindrical objects. The extremely simplified object approximation with ellipses became insufficient as soon as the manipulation of objects with more complex shape was required. Also, the results were only possible with the hybrid visual servoing by inclusion of the smart tray into the control structure.

Additional restrictions arose through the visual servoing loop. In particular, both the object to be manipulated and the robot gripper had to be visible continuously in both 2D images. This condition limited the work space of the robot arm and consequently of the whole system to a rather small area in front of the stereo camera system and also to a quite artificial robot configuration.

In order to resolve the restrictions of vision-based control in FRIEND I and to enable reliable robot operation in an unstructured environment and to enhance manipulability, further development is now focused on improvement of FRIEND's vision capabilities. In particular, methods for robust 3D object reconstruction are researched and developed.

2.3 Vision-Based Control of 7 DOF Robot Arm in FRIEND

The development of the robust 3D object reconstruction methods which provide reliable information on the pose of objects to be manipulated was performed in parallel to the development and implementation of a precise light weight robot arm with 7 DOF. The built-in sensors of the robot provide precise information on the position of the gripper and of all joints. Furthermore, currently available calibration algorithms allow a precise determination of the camera coordinate system in relation to the chosen world coordinate system. The combination of these elements led to our decision to implement a position-based visual control structure into FRIEND as shown in Fig. 5. This choice of visual robot arm control also allows easy integration of collision free path planning. The main modules of this vision-based control structure are described briefly in the following.

Robust 3D Real-World Object Reconstruction in FRIEND

3D object reconstruction in FRIEND is based on novel method for 3D modelling of real-world objects which enables autonomous object manipulation in an unstructured environment and defines 3D object models reliably without a priori knowledge on object geometry. Crucial for 3D reconstruction is the robust object segmentation in stereo images of the FRIEND vision system. The FRIEND segmentation method combines disparity based segmentation with robust closed-loop colour region based segmentation. The disparity map segmentation leads to the definition of object region of interest (ROI) and assures robustness against cluttered scenes.

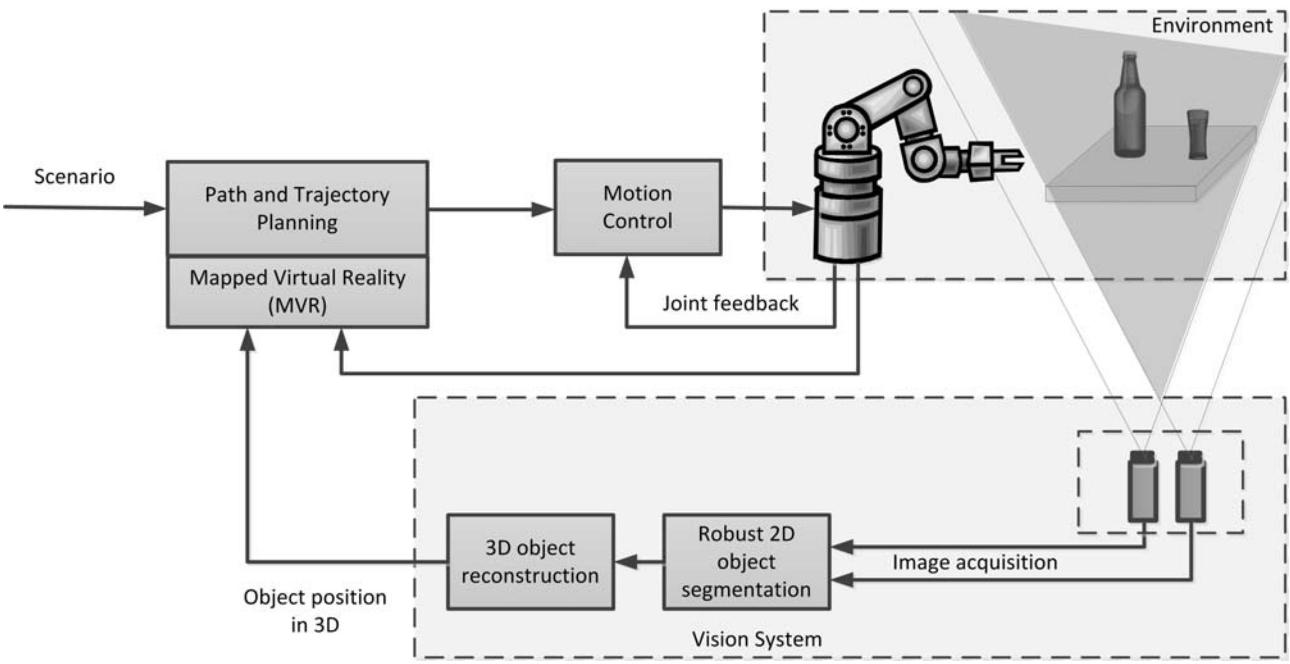

Figure 5 Position-based control with internal joint control of the robot.

In the determined ROI the objects are segmented from their background, using closed-loop colour region based segmentation which assures robustness against variable illumination. The inclusion of feedback control at segmentation level was proposed by authors [4; 9]. The main idea behind feedback structures in image processing is to change image processing parameters in a closed-loop manner until the desired processing quality is achieved independently of external influences such as variable illumination. The FRIEND 2D object segmentation is described in detail in [10]. In this paper it is just mentioned briefly to stress its main function, providing reliable input to the 3D object reconstruction module which estimates the object geometry and pose. It is also important to stress that in contrast to state-of-the-art methods [11] the FRIEND object recognition method is independent of textured features and can recognize objects with texture as well as texture less objects so that it enables reconstructions of real-world objects in the robot's unstructured environment.

The real-world object modeling implemented in FRIEND is based on 3D contours which are based on good continuation of the local feature primitives such as edges. As such, 3D contours provide geometric and shape information on the object. The details of robust segmentation of different objects from the "serving a drink" scenario can be found in [10]. In the following, as an example, the segmented image of a "Bonaqua" water bottle from the image scenes in Fig. 2 is considered. The segmented bottle object in the left stereo image is given in Fig. 6a.

In the process of computation of the 3D object contour, first multiple convex hull fitting is performed. Gaussian kernel based contour smoothing [12] is applied on the segmented object boundaries to overcome small perturbations due to noise and to get a smooth object curvature. As the change in gradient Δx of object curvature represents the local shape of the object, 2D dominant shape points are defined on the object boundary according to:

$$\begin{aligned} &\Delta x = x_i - x_{i-1}\,; \quad i = 1\ldots b \\ &k = k+1\,, \quad \text{if abs}(\Delta x > 0)\,, \end{aligned} \tag{4}$$

where k is the number of dominant shape points, x_i, x_{i-1} are horizontal pixel coordinates of the neighboring boundary points in the image and b is the number of 2D boundary points on one side of the objects symmetry axis. Due to the presence of lateral symmetry in the object, the dominant points are computed only on one side of the objects symmetry axes as illustrated in Fig. 6a.

A convex hull between two neighboring dominant shape points results in a 'slice' of the segmented object. In general, the 2D robustly segmented object image is divided into k slices based on the number of dominant points as seen in Fig. 6b.

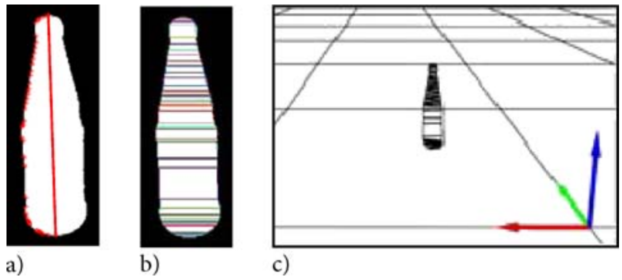

Figure 6 Multiple hull fitting procedure. (**a**) Dominant shape points on one side of the object symmetry axis. (**b**) Multiple hulls fitted on the 2D robust segmented left stereo image. (**c**) Computed 3D contour of the object.

To compute the 3D object contour, it is necessary to solve the stereo correspondence problem. The stereo correspondences of the determined dominant shape points in the left stereo image are determined by searching along the epipolar line in the right stereo object image [13]. The obtained stereo correspondences are represented as

$$\begin{aligned} &p_{L_i} = (x_{L_i}, y_{L_i})\,, \\ &p_{R_i} = (x_{R_i}, y_{R_i})\,, \quad i = 1\ldots k \end{aligned} \tag{5}$$

where p_{L_i}, p_{R_i} are the obtained stereo correspondent points in the left and right object image respectively and k is the number of dominant shape points. Due to both, different perspective views of the stereo cameras and external influences like noise, the obtained stereo correspondences may not satisfy the epipolar constraint (6).

$$p_{R_i}^T \mathbf{F} p_{L_i} = 0\,, \tag{6}$$

where $\mathbf{F}$ is the fundamental matrix which relates the corresponding points in the stereo images. To overcome this problem, the optimal stereo correspondences are computed by minimizing the sum squared distances between the obtained image points and the back-projected point on the epipolar line satisfying the epipolar constraint [10]

$$(\hat{p}_{L_i}, \hat{p}_{R_i}) = \operatorname{argmin}\left\{d_L(p_{L_i}, l)^2 + d_R(p_{R_i}, l')^2\right\} \tag{7}$$

where $\hat{p}_{L_i}$, $\hat{p}_{R_i}$ are the optimal stereo correspondences in the left and right object image, l, l' are the epipolar lines in the left and right object images respectively, d_L, d_R are the Euclidean distance between the image point and the point on the epipolar line. Minimization procedure of (7) using first order approximation [13] yields optimal correspondences which are the maximum likelihood estimates for the true image point correspondences. Hence, the obtained optimal correspondence points are robust against noise and outliers. With the known camera matrices and the optimal correspondences for the dominant shape points, the 3D point is found by the intersection of the two projection lines in the 3D space using the linear stereo triangulation procedure described in [14]. The obtained 3D boundary points for each slice in the multiple hulls are approximated by fitting a 3D contour.

The resulting multi modal 3D contour of the bottle in Fig. 6b is shown in Fig. 6c. As obvious, the resulting 3D object contour allows generation of 3D object model that fits to the actual shape of the object.

In [10] 3D modelling of cylinder as well as of cuboid objects starting from computed 3D contours was explained. In this paper, the 3D modeling of real-world objects from the robot's environment is illustrated by the modelling of a cylinder object such as the above considered bottle object. In order to model the full 3D view of a cylinder object, both the radius of the brim circle and the object height is to be determined with high accuracy from the 3D object contour. Due to the viewing configurations of the stereo camera, only one face of the object is imaged. For the "serving a drink" scenario the assumption is made that objects are symmetrical around their vertical axis. The vertical axis of the object is calculated using the surface normal vector information of the 3D boundary points in 3D contour. The height and radius of each 3D contour is estimated to model its equivalent cylindrical structure about the vertical axis of the object. The shortest distance between the two horizontal 3D lines in the contour gives the 3D contour height while the shortest distance between the two vertical 3D lines gives the contour width. The result of 3D modelling of the bottle object which is shown in Fig. 2b (bottle in the image of table scene) is given in Fig. 7a.

The generated 3D models of objects to be manipulated are used further to update the workspace representation, which is implemented into FRIEND as a Mapped Virtual Reality (MVR) [15], see Fig. 7b. Beside 3D models of objects to be manipulated, 3D reconstruction of so-called "container" objects is used for MVR adaptation. "Containers" are objects in the FRIEND environment such as the fridge, table, microwave oven and book shelf. Hence, "containers" are the objects in which/on which the objects to be manipulated are placed. Bearing in mind that the container objects in the FRIEND environment are a permanent feature of the scenarios, the SIFT method [16] is used for their recognition. This method uses an artificial marker as a model image and during on-line system operation the SIFT algorithm searches for the model image in the scene through a matching based

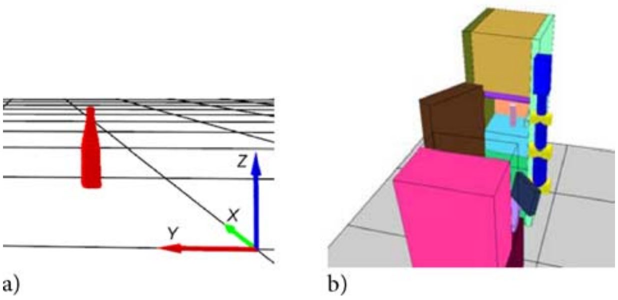

Figure 7 (**a**) 3D model of the reconstructed bottle object from the image of table scene of "serving a drink" scenario (**b**) Mapped Virtual Reality (MVR) representation of the robot environment.

algorithm. Once the model image has been detected, its pose (position and orientation in 3D space) can be reconstructed using model based matching. Knowing the position of the model image placed on/in a container (e. g. in the fridge), the container pose can further be reconstructed. It was necessary to enhance the real time behaviour of SIFT before it could be used in FRIEND. The details of the implemented "VF-SIFT" method can be found in [17]. In this paper, the correct 3D reconstruction of containers achieved by SIFT model based method is used as the input to MVR.

MVR uses a simplified representation of the robot arm and of real-world objects in its environment. It is used for online monitoring of the robot workspace and prediction of possible collisions, which is required for safe object manipulation and trajectory planning [15]. Before a reliable object recognition was available, domain specific knowledge was used to describe the environment in MVR. With the availability of the above described object recognition, the MVR is not completely abandoned but used as a redundant source of information. The MVR is initialized with domain specific knowledge like a priory known 3D-models of the permanent objects in the scenarios (such as "container" objects). The MVR model is then continuously updated and extended with the information on location of dynamic objects in the robot workspace using the online generated 3D-models of objects to be manipulated and also 3D data on "containers" pose.

Path and Trajectory Planning

The module of visual control structure in Fig. 5 named "Path and Trajectory Planning" generates the collision-free path. The input of the trajectory planning is derived from different nets (a hierarchy of AND/OR nets and Petri nets) [18] to which we refer here as the scenario. The scenario contains primarily information on the objects to be manipulated and the task that the manipulator should accomplish. For example the "serve a drink" scenario describes among others that the recognised object "bottle" should be grasped and be taken out of the fridge. The scenario is used by the scheduler to request a series of manipulative operations which are responsible for executing manipulator tasks. For each operation, a trajectory must be constructed and executed by the robot arm. The inputs for the trajectory calculation are the target object and MVR model of the scene. The procedure includes three phases:

- Calculation of the goal configuration
- Collision free path generation
- Trajectory calculation.

Calculation of the goal configuration: As already mentioned, the location of the object is provided by the vision system and the 3D model is added to the MVR. Using this location, a Tool Center Point (TCP) frame is calculated and a goal configuration is computed. The latter is selected from an on line calculated set of possible inverse kinematic solutions which correspond to the same TCP frame. Due to the 7 DOF in theory, an infinite number of solutions exist, but only a limited set is calculated. The criteria for selecting one out of the large number of possible configurations are the maximum distance between manipulator and obstacles and the minimum difference between a solution and the start configuration of the robot. Using the symbols {W} as the world, {G} as gripper and {O} object's coordinate system, the transformation between the world and final end effector TCP frame {G′} can be calculated as follows [19].

$$\mathbf{T}^{W}_{G'} = \mathbf{T}^{G}_{O} \cdot \left(\mathbf{T}^{W}_{O} \cdot \mathbf{T}^{O}_{sample}\right)^{-1} \tag{8}$$

Figure 8 illustrates each coordinate system in the equation (8). The $\mathbf{T}^{O}_{sample}$ is a sample frame calculated from a random position and rotation that the target object can have in the object's coordinate system. For instance, a bottle can be rotated about the Y axes of the $\mathbf{T}^{W}_{O}$ frame.

Collision Free path generation: A new path planning approach is implemented into the current FRIEND generation. The inputs for the path planning are the goal configuration and the actual MVR scene. The planner is called *CellBiRRT* [20]. It is based on the concept of Rapidly exploring Random Trees (RRT). The main feature of this approach is the ability to solve path planning tasks while, at the same time, it can handle additional position and orientation constraints of the end-effector. This is done by dividing the Cartesian space into cells and selecting an appropriate one as a basis for generating random configurations. One of these random configurations is used later in order to expand the trees randomly.

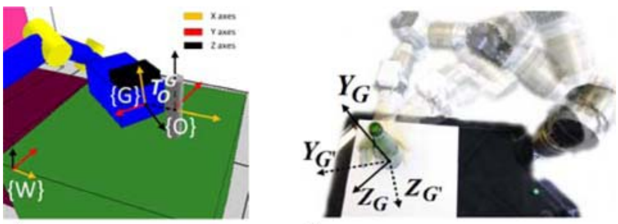

Figure 8 (**a**) MVR representation of grasping scene with related coordinate systems (**b**) Example of different inverse kinematics solution for different $\mathbf{T}^{O}_{sample}$.

Figure 9a, b presents the cells and how the configuration is chosen. Figure 9c, d shows the path generation result for a part of the "serve a drink" scenario namely grasping the bottle in a fridge and placing it on a serve table.

A path planning algorithm requires an efficient method for computing distances and detecting possible collisions of objects. A configuration is in collision if the overall minimum distance between robot arm and obstacles is less than a tolerance limit which is equal to the scene 3D reconstruction's error of the vision system. The collision detection method used in FRIEND is presented in [21]. It distinguishes near and far obstacles and reduces dynamically the number of samples in a segment between two configurations that have to be checked for collision. It is a combination of oversized Oriented Bounding Boxes (OBBs) and an algorithm for calculating the minimum distance between two convex polyhedrals. The algorithm for calculating the minimum distances is the Gilbert-Johnson-Keerthi (GJK) algorithm [22].

A *trajectory calculation* step is included to enhance the path generated in previous steps in order to make the robot movement smoother. The resulting trajectory is then provided to the robot controller in order to move the robot arm along it.

Performance Evaluation

In order to evaluate the performance of the vision-based control in the current FRIEND system, experiments were conducted in which the robustness of the described 3D object reconstruction as a key factor for reliable manipulator control was assessed. 100 test images containing objects that are commonly found in both phases of the "serve a drink" ADL scenario were taken in different illumination conditions. The objects were imaged under both daylight and artificial lighting conditions. The lighting conditions were varied from 50 to 500 lx. This range of illumination corresponds to a variation of the light intensity from a family living room (50 lx) to the standard lighting level of an office (500 lx).

The objects in each test image were segmented using the proposed segmentation method which combines disparity map segmentation with closed-loop colour region segmentation. The segmented object images were used as reliable inputs to the 3D object reconstruction module

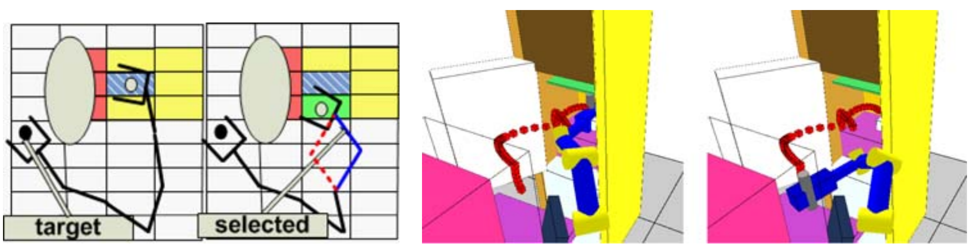

Figure 9 (**a, b**) Examples of cell selection. The cell with minimum distance to a target TCP is selected. (**c, d**) Path generation for placing bottle on the table. Start (**c**) and goal (**d**) configurations are presented.

Table 2 Experimental results for cylinder object reconstruction.

Objects		Height [m]	X [m]	Y [m]	Z [m]	Volume Fraction V_f [%]
Sauce Jar in fridge	Ground Truth	0.110	0.183	0.533	0.095	100
	Reconstruction in 500 lx experiment	0.111	0.181	0.530	0.092	98.32
	Mean for 15 trials	0.106	0.176	0.532	0.092	96.48
	σ	0.003	0.004	0.005	0.001	1.490
Bottle of "bonaqua" water in fridge and on the table	Ground Truth	0.285	0.355	0.375	0.340	100
	Reconstruction in 500 lx experiment	0.287	0.357	0.378	0.337	98.42
	Mean for 15 trials	0.286	0.353	0.373	0.338	96.39
	σ	0.001	0.003	0.004	0.002	3.44
Glass on the table	Ground Truth	0.125	0.300	0.280	0.125	100
	Reconstruction in 500 lx experiment	0.124	0.294	0.277	0.124	98.17
	Mean for 15 trials	0.123	0.296	0.275	0.123	97.06
	σ	0.002	0.003	0.004	0.001	1.14

where the 3D coordinates X, Y and Z of the object grasping point were calculated. The automatically calculated 3D objects coordinates were compared with the ground truth position errors obtained in off-line experiments using a high precision laser range finder. Also geometrical characteristics such as height, and radius for cylindrical objects were computed from the extracted 3D contour and the computed results were compared to ground truth data. In order to compute the accuracy of the generated 3D model, the volume of the reconstructed object was compared with the actual volume of the considered object. The measure of reconstructed volume fraction (V_f) is defined as:

$$V_f = \frac{Volume_{reconstruted}}{Volume_{actual}} \times 100[\%] \quad (9)$$

where $Volume_{actual}$ represents off-line determined ground truth.

The required precision of 3D reconstruction to successfully grasp various objects without toppling them was evaluated in off-line experiments for the considered ADL scenario. The largest position error in each direction for which the object could still be successfully grasped was calculated for various real-world objects. The mean of error tolerance for the different dimensions over 50 trials is presented in Table 3.

The experimental results for representatives of cylindrical objects which are usually present in both scenes (in the refrigerator and on the table) of the "serve a drink" scenario are shown in Table 2. The results of object reconstructions in one single experiment are given together with statistical descriptors of a number of experiments. Reconstruction in the 500 lx experiment refers to the obtained object reconstruction results under office lighting conditions (500 lx), while the statistical results are obtained from 15 trials under variable illumination conditions. From the presented experimental results, it is observed that the error in 3D reconstruction is within the tolerance limits given in Table 3 and, hence, the objects can be dexterously grasped by the manipulator.

Table 3 Error tolerance levels X_T, Y_T and Z_T in X, Y and Z direction for 3D reconstruction accuracy for object grasping.

X_T [m]	Y_T [m]	Z_T [m]
0.015	0.020	0.010

3 Conclusions

In this paper, the experiences of IAT in visual robot control in different generations of the FRIEND system have been presented and also the key technological developments are presented which led to the implementations in the newest FRIEND generation.

During the development of FRIEND, a pragmatic balance has been sought and achieved between, on the one hand, developing innovative solutions based on the latest research and available technology, and on the other hand, keeping to a time frame for each generation that enabled an acceptable rate of progress towards the objective of using the system in a real world based application. Presently FRIEND is being prepared for the support of a completely paralysed person who will be reintegrated in professional life and who will work in the University Library (SUUB) as a librarian who retro-catalogues books into the data base. All manipulative skills necessary to handle the books at that working place will be carried out by FRIEND.

Acknowledgements

We would like to mention that some very important IAT achievements that played an important role in the constant improvement of the assistive robot FRIEND could not be described here due to limited space. Some of these contributions are R&D of the event based control con-

cept, the basic software system MASSiVE [23], the safety concept and the Human Machine Interface as well as different advanced image processing techniques.

We would like to thank here all IAT researchers in these fields, particularly Dr. C. Martens, Dr. O. Prenzel and Dipl.-Ing. E. Wendland for their very important contributions and project leadership over many years.

Many other people from outside IAT contributed and still contribute to the R&D of FRIEND and we would like to mention especially Dr. Spranger and his team from the Neurologisches Rehabilitationszentrum Friedehorst (NRZ) and Mr. Ahlborn from the SUUB Bremen.

Last but not least we would like to mention the University of Bremen, the Federal Ministry of Education and Research (BMBF) and the Integration Agency of Bremen[3] for their important financial support of the research

References

[1] S. Hutchinson, G. D. Hager, P. Corke: A tutorial on visual servo control. – In: IEEE Transactions on Robotics and Automation **12**, 651–670 (1996).

[2] F. Chaumette, S. Hutchinson: Visual servo control I: Basic approaches. – In: IEEE Robotics and Automation Magazine **13**(4), 82–90 (2006).

[3] C. Martens, O. Prenzel, A. Gräser: The Rehabilitation Robots FRIEND-I & II: Daily Life Independency through Semi-Autonomous Task-Execution. – In: Rehabilitation Robotics, I-Tech Education Publishing, Vienna, Austria 2007.

[4] D. Ristić: Feedback Structures in Image Processing. Shaker Verlag 2007.

[5] R. B. E. Röme, H. J. A. Stuyt, A. Peters: Cost-Savings and Economic Benefits due to the Assistive Robotic Manipulator (ARM). – In: Proc. of 9th International Conference on Rehabilitation Robotics (ICORR 2005), 461–464(2005).

[6] I. Volosyak, O. Kouzmitcheva, D. Ristić, A. Gräser: Improvement of Visual Perceptual Capabilities by Feedback Structures for Robotic System FRIEND. – In: IEEE Transactions on System, Man, and Cybernetics Part C (February 2005), 66–74, (2005).

[7] O. Kouzmitcheva, C. Martens, A. Pape, H. She, I. Volosyak, A. Gräser: FRIEND – An Intelligent Assistant in Daily Life. In: Advances in Rehabilitation Robotics: Human-friendly Technologies on Movement Assistance and Restoration for People with Disabilities, Z. Z. Bien, D. Stefanov (Eds.), Springer 2004.

[8] O. Kouzmitcheva, A. Gräser: Rehabilitation System FRIEND: A Few Steps towards User Friendly and Reliable System. – In: Proc. of the International Conference on Intelligent Manipulation and Grasping IMG04, Genova, Italy, 208–213 (2004).

[9] S. M. Grigorescu, D. Ristić-Durrant, A. Gräser: RObust machine VIsion for Service robotic system FRIEND. – In: Proc. of the 2009 IEEE Int. Conf. on Intelligent RObots and Systems, St. Louis, USA, (2009).

[10] S. K. Natarajan, D. Ristić-Durrant, A. Leu, A. Gräser: Robust stereo-vision based 3D modelling of real-world objects for assistive robotic applications. – In: Proc. of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011), San Francisco, California, (2011).

[11] K. Welke, T. Asafour, R. Dillmann: Active multi view object search on a humanoid head. – In: Proc. of the 2009 IEEE Int. Conf. on Robotics and Automation, Kobe, Japan, (2009).

[12] K. Kpalma, J. Ronsin: Contour smoothing for non-occluded planar shape description. – In: Proc. of the IEEE Int. Conference on Image Processing, 381–384, (2005).

[13] R. Hartley, A. Zisserman: Multiple View Geometry in Computer Vision. Cambridge University Press 2004.

[14] S. M. Grigorescu, S. K. Natarajan, D. Mronga, A. Gräser: Robust feature extraction for 3D reconstruction of boundary segmented objects in a robotic library scenario. – In: Proc. of the 2010 IEEE Int. Conf. on Intelligent RObots and Systems, Taipei, Taiwan, (2010).

[15] J. Feuser, O. Ivlev, A. Gräser: Collision Prevention for Rehabilitation Robots with Mapped Virtual Reality. – In: Proc. of the IEEE 9th International Conference on Rehabilitation Robotics (ICORR 2005), Chicago, IL, USA, (2005).

[16] D. G. Lowe: Distinctive image features from scale-invariant keypoints. – In: Int. Journal of Computer Vision **60**(2), (2004).

[17] F. Alhwarin, D. Ristić-Durrant, A. Gräser: VF-SIFT: Very Fast SIFT Feature Matching. – In: Pattern Recognition Lecture Notes in Computer Science, Springer Verlag Berlin, Heidelberg, Volume 6376/2010, 222–231, DOI: 10.1007/978-3-642-15986-2_23, (2010).

[18] T. Cao, A. C. Sanderson: AND/OR Net Representation for Robotic Task Sequence Planning. – In: IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews **28**(2), (1998).

[19] F. Schwarzer, M. Saha, J. C. Latombe: Adaptive dynamic collision checking for single and multiple articulated robots in complex environment. – In: IEEE Transactions on Robotics **21**(3), (2005).

[20] C. Fragkopoulos, A. Gräser: A RRT based path planning algorithm for Rehabilitation robots. – In: ISR/Robotics, Munich, (2010).

[21] C. Fragkopoulos, A. Gräser: Dynamic efficient collision checking method of robot arm paths in configuration space. – In: Proc. of the IEEE/ASME International Conference on Advance Intelligent Mechatronics, (2011).

[22] S. Cameron: Enhancing gjk: Computing minimum and penetration distances between convex polyhedra. – In: Proc. of the in IEEE Int. Conference on Robotics and Automation, 22–24, (1997).

[23] C. Martens, O. Prenzel, J. Feuser, A. Gräser: MASSiVE: Multi-Layer Architecture for Semi-Atunomous Service Robots with Verified Task Execution. – In: In Proc. of the 10th International Conference on Optimization of Electrical and Electronic Equipments – OPTIM 2006 (Brasov, Romania: Transylvania University Press, Brasov) 3, 107–112, (2006).

Received: February 9, 2012

Prof. Dr. Axel Gräser received the diploma in electrical engineering from the University of Karlsruhe, Germany, in 1976 and the Ph.D. degree in control theory from the TH Darmstadt, Germany, in 1982. Since 1994, he has been the Director of the Institute of Automation, University of Bremen, and the Head of the Department of Robotics and Process Automation. His research interests include service robotics, brain robot interface, digital image processing and augmented reality.

Address: Institute of Automation, University of Bremen, Otto-Hahn-Allee, NW1, 28359 Bremen, e-mail: ag@iat.uni-bremen.de

Dr. Olena Kuzmicheva received the diploma in automation and control from the Dnepropetrovsk State University, Ukraine, in 1999 and Ph.D. degree in Electrotechnical Engineering from the University of Bremen, Germany, in 2009. After her Ph.D. studies, she has been a postdoctoral researcher and lecturer at the Institute of Automation, University of Bremen, and Friedrich-Wilhelm-Bessel-Institute Research Society in Bremen, Germany. Here research interests are service robotics, robotics for risky interventions and human machine interfaces.

Address: Institute of Automation, University of Bremen, Otto-Hahn-Allee, NW1, 28359 Bremen, e-mail: olena@iat.uni-bremen.de

[3] Integration Agency of Bremen is a part of the Pension Office in Bremen, assigned to the Department of the Bremen Senator for Profession, Women, Health, Youth and Social Affairs

Dr. Danijela Ristić-Durrant received the M.Sc. degree in automatic control and robotics from the Faculty of Mechanical Engineering, University of Niš, Serbia, in 1998 and Ph.D. degree in Electrotechnical Engineering from the University of Bremen, Germany, in 2007. After her Ph.D. studies at the University of Bremen, she has been a postdoctoral researcher and lecturer at the IAT. Her research interest is interdisciplinary research in control theory and computer vision within the research on robust robot vision and vision-based human gait analysis.

Address: Institute of Automation, University of Bremen, Otto-Hahn-Allee, NW1, 28359 Bremen, e-mail: ristic@iat.uni-bremen.de

Saravana K. Natarajan received the BE degree in Electrical Engineering from Bharathiar University, India in 2004 and M.Sc. degree in Information and Automation from Universität Bremen, Germany in 2008. He is currently a Ph.D. student at Institute of Automation, University of Bremen. His research interests include computer vision and 3D reconstruction of real world objects.

Address: Institute of Automation, University of Bremen, Otto-Hahn-Allee, NW1, 28359 Bremen, e-mail: snatarajan@iat.uni-bremen.de

Christos Fragkopoulos received the Diploma degree in Physics from the Aristoteles University, Thessaloniki, Greece, in 2004, and the M.Sc. degree in Information and Automation Engineering from the University of Bremen, Bremen, Germany in 2007. Since 2008 he has been working toward the Ph.D. degree at the Institute of Automation, University of Bremen, Bremen, Germany in the field of manipulator motion planning and control. His main areas of interest are robotics and artificial intelligence.

Address: Institute of Automation, University of Bremen, Otto-Hahn-Allee, NW1, 28359 Bremen, e-mail: cfragkopoulos@iat.uni-bremen.de