ACTIVE INSPECTION AND HANDLING OF UNKNOWN OBJECTS USING AN AUTONOMOUS HAND-ARM-EYE SYSTEM

M. SEITZ, U. HOLESCHAK and K. KLEINMANN *

**Darmstadt University of Technology, Control Systems Theory and Robotics Department, Landgraf-Georg-Str. 4, D-64283 Darmstadt, Germany*

Abstract. *Future applications to service robots require autonomous handling of various objects appearing in every-day-life. This paper describes a hand-arm-eye system inspecting arbitrarily shaped objects from multiple views. The vision system is based on edge extraction merging edges in contour images acquired from the scene in order to extract favourable grasp features. Contour segments favoured for being contacted by the fingers of the employed dextrous gripper are localized by 3D active inspection. Combining characteristic grasp features selected from multiple views, optimum contact positions are planned allowing to enclose a grip or a handle by the fingers. In particular, problems due to the interaction between different elements of the integrated hand-arm-eye system will be discussed. Experiences from experimental results allow to evaluate the realizability and expense of executing complex robotic tasks and to examine the possibilities and limits of the presented system.*

Key Words. hand-arm-eye system, 3D active vision, contour image enhancement, grasp planning

1 INTRODUCTION

Executing complex robotic tasks including grasping and manipulation requires a combination of flexible actuators, intelligent sensors and adequate object information processing. Due to the uncertainty in the real world intelligent robotic systems need rich sensory information to understand the state of their environment. Additionally, highly redundant robots are necessary to provide sufficient flexibility for dexterous manipulations. While the integration of multiple sensors into a robotic system increases the system autonomy, it also introduces some problems due to the different modules to be coordinated and the complexity of acquired data to be condensed efficiently depending on the specific context. Though there are many interesting approaches examining the subsystems of such complex robotic systems towards their integration, e.g. (Bard *et al.*, 1994), there is a lack of experimental results applying to those integrated systems.

Therefore, in this work a vision integrated, highly redundant robotic system is presented able to perform complex tasks like grasping and manipulating a-priori unknown, volumetric objects as they would have to be handled by service robots in the household or in the field of nursing. The basic modules of the system are a 3D vision assisted grasp planner using highly condensed context information, a rotating camera and a three-fingered dexterous gripper both mounted on a six-axes puma-type robot arm as shown in fig. 1. This hand-arm-eye system supplies seven degrees of freedom for observation and fifteen for manipulation. The employed experimental gripper was developed and built at the Darmstadt University of Technology (Paetsch and Kaneko, 1990). The flexibility of camera and gripper allows entire object inspection as well as performing various grasps and manipulations.

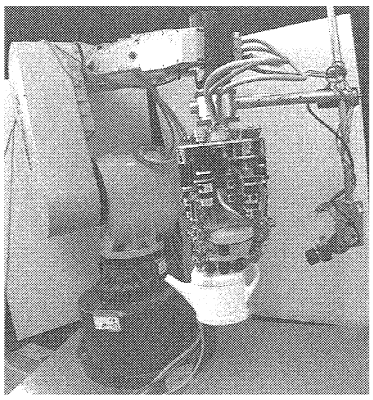

Figure 1: *The hand-arm-eye system consists of a multi-fingered gripper and a mobile camera both mounted on a robot arm. Guided by vision assisted motion planning the system handles a watering can.*

The main task of such a hand-arm-eye system is to inspect and grasp an object. Planning optimal grasps for different kind of grippers has been examined from the theoretical point of view e.g. by (Ferrari and Canny, 1992), (Park and Starr, 1992),

assuming that the object representation is given. In order to avoid a limitation of the work space by a static field of vision, a mobile camera carried by the robot is used for autonomous perception of adequate object information. Most of these arm-eye systems employ parallel-jaw grippers (Taylor *et al.*, 1994) for which stable grasps can easily be planned. But flexible manipulations require dexterous multifingered grippers. For a secure determination of stable fingertip contact points, a complex shape analysis is necessary (Seitz and Kraft, 1994). Recently, our dexterous hand-arm-eye systems has been employed to grasp planar objects inspected only from the top view (Weigl and Seitz, 1994). But relatively large and complex objects like the watering can of fig. 1 have rarely been examined by those approaches, because an entire inspection from multiple views is necessary for planning autonomous gripper approach and grasping. For the sake of higher security and stability it is sometimes desirable to contact those objects not only by the fingertips but by finger links and the palm, also. During this contact phase inner link force/torque (Kaneko and Honkawa, 1994) or tactile sensor information (Reznik and Lumelski, 1994) can perceive local object information and an optimum grasp configuration can be found. Beside this local analysis of suitable contact forces, global object information is required, additionally, in order to plan which object part can be contacted by the links and from which direction the gripper has to approach.

In this paper an autonomous hand-arm-eye system is presented inspecting arbitrarily shaped objects from multiple views in order to plan stable fingertip or power grasps. The performance of those grasps by the Darmstadt hand has been studied already (Paetsch *et al.*, 1993), assuming the object to be in an optimum well-known position between the before preshaped fingers. In contrast to this approach the major goal of this work is to plan the motion of the hand-arm system from an arbitrary robot position to the preshaping positions of the fingers by the use of vision. Sample experiments have been examined in order to evaluate the limits as well as the applicability and reliability of the system and to discuss practical problems due to the interaction between different elements of the integrated hand-arm-eye system.

In the following, the vision-based extraction of suitable grasp features is explained as well as the interaction between robot and camera for optimum object inspection. After that, the monocular information acquired from different views is composed to a 3D object representation used for grasp planning and performance. Finally, the feasability of our system and problems revealed by experiments will be discussed.

2 GRASP FEATURE EXTRACTION

In contrast to planar binary image analysis sufficient for many applications using a hand-arm-eye system, e.g vision assisted disassembly of a video camera recorder (Weigl and Seitz, 1994), contour image analysis is required for inspection of volumetric objects. Contour images provide a more detailed object representation in order to detect interior edges describing grips or handles not visible in a silhouette. Furthermore, edges are robust features which can be recognized and tracked in an image sequence.

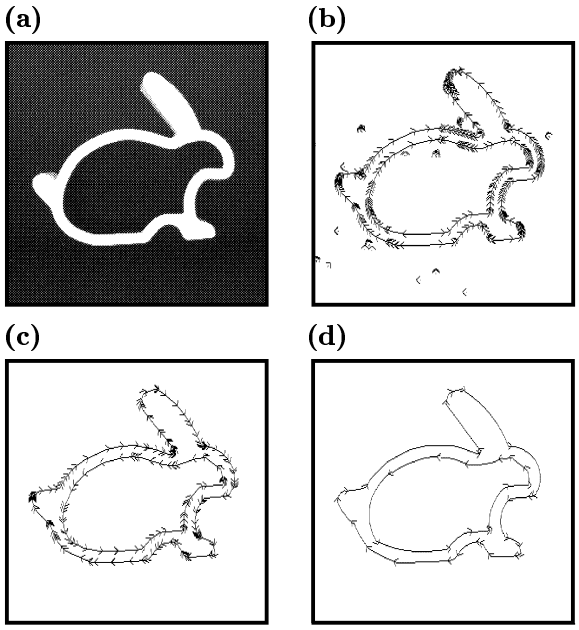

Figure 2: *The steps of image enhancement result in an extreme data reduction. At first the grey-level image (a) is transformed to a vectorized contour image (b). The vectors indicated by arrows (>) are filtered (c) and merged to contour circle segments (d).*

To explain the single steps of the image processing part of our system, which is hierarchically structured and supported by a real-time edge detector, the planar baking tin shown in fig. 2 is regarded as sample object. At first the grey-level image (fig. 2a) is reduced to a contour image by edge extraction. A professional edge extraction and vectorization hardware is used for acquisition of a highly condensed but disturbed symbolic representation of the object contour. Figure 2b is an example for a contour image disturbed by noise and discontinuous or interrupted edges. Moreover, the topology of neighbouring contours is not reliably determined.

To enhance this image, short vectors due to noise are deleted, interrupted vectors are connected and little discontinuities in the contour are smoothed by using compensating straight lines. Furthermore, the neighbourhood relations between contours are corrected and the direction of closed contours indicated by the arrows in fig. 2b-d is homogenized. This efficient contour filtering results in a data reduction of 50% (fig. 2c) compared to the initially acquired contour image (fig. 2b).

For efficient grasp feature extraction a more condensed contour representation is necessary. Therefore, the vectors are approximated by circle segments of variable curvature. Following the vectors of the enhanced contour, the curvature of a segment is estimated by local averaging the change of vector directions along the contour. In each step it has to be decided if the actually examined vector is a member of the merged segment or of a new segment. To evaluate this, the error between the merged vectors and the approximated circle segment is determined. If its value increases a static threshold, a new segment to be approximated is started with the actually examined vector.

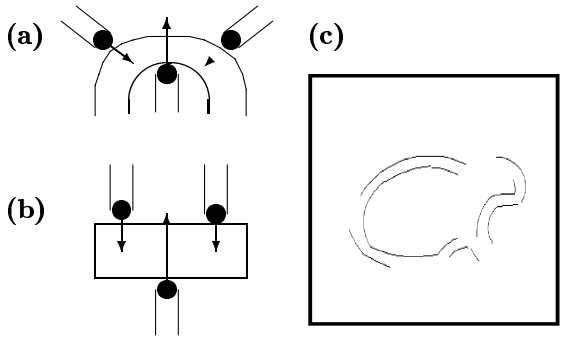

Figure 3: *Basic grasp configurations on parallel curved (a) or parallel straight contour parts (b). Accordingly, parallel circle segments (c) are extracted as favourable grasp features from the contour of fig. 2.*

This approximation of vectors by curved contour segments results in a highly condensed contour representation reducing the data to 10% of the initial amount. The expense of computation time ($< 1sec$ for filtering the sample object) can be neglected with respect to the expense for grasp planning and the time needed for robot motions. The remaining points in fig. 2d separate contour segments approximated by circle segments.

To evaluate which contour segments describes suitable contact regions for the fingers of a three-fingered gripper, two typical grasp configurations are regarded. Either the fingers are arranged in approximately 120^o angles apart to each other (fig. 3a) or in parallel to each other (fig. 3b). Which configuration is preferred, depends on the specific context characterized by the geometry of the employed gripper, the given task and the extracted features of the object.

Therefore, those contour segments are searched in the filtered contour image (e.g. fig. 2d), to which the favoured grasp configurations can be applied. Basically, contour segments with almost the same curvature located in parallel to each other are regarded as favourable grasp features. By reduction of the entire image data (fig. 2) to those parallel circle segments as shown in fig. 3c, almost every favourable contact region on an object can be modelled. The extracted parallel segments are merged to feature groups from which the 3D grasp planning (see section 4) determines the most favourable contact points.

In order to provide a maximum contact area, the selected grasp features are expanded as far as possible again by merging neighboured contour segments. Using the approximated circle segment, the contour of a grasp feature can be extrapolated in order to detect other contour segments of almost the same curvature in its close neighborhood. If those segments can be found, they are combined by connecting the discontinuities. An example for this, is the connection of parallel lines describing handle and spout of the watering can shown in fig.5a.

After extraction of favourable grasp features from the acquired camera image, the location of these features has to be determined in world coordinates. Therefore, the robot moves the camera around the object in order to reconstruct the feature shape from motion.

3 ACTIVE OBJECT INSPECTION

In this approach robot motions are not only performed for grasping and manipulating objects but in particular for fine object inspection. In the following autonomous robot motion planning due to the perceived object information is explained starting with an automatic object zooming followed by camera navigation for 3D feature localization.

3.1 Automatic Object Zooming

Starting object inspection, an arbitrary object position in the field of vision has to be assumed. For fine object inspection the object is zoomed to an optimum representation in the image. Moving the camera perpendicularly above the scene, the position and size of the object can be estimated determining the displacement of the object enclosing frame in the image (cf. fig. 4a). For a quick determination of the coarse object orientation, the smallest frame enclosing the object is searched by trial of three differently rotated frames. The principal axis of the resulting frame is approximately the same as that of the object itself. Due to the determined object location, orientation and size a camera motion is performed in order to align the object to a desired image representation. But the desired representation can hardly be achieved in one step because unprecise camera calibration results in inaccurate motion planning. Therefore, object pose and size are tracked in a control loop sketched in fig. 4c, until the desired alignment is achieved (fig. 4b). Although there are some approaches in literature, e.g. (Yoshimi and Allen, 1994), performing similar object alignment procedures without camera calibration, in this approach the use of a coarse camera calibration allows quick convergence and avoids instability of the control loop behaviour. So, the main motion is performed in the first step and the control loop is only needed for small corrections.

The acquired optimum object representation in the image allows an object fine inspection in order to extract and localize the suitable grasp features.

3.2 Grasp Feature Localization

In order to grasp an object on a grasp feature extracted from the acquired image (as described in section 2), the feature has to be localized. For 3D localization two contour images acquired e.g. from the top view of a watering can (fig. 5a) are analyzed. Corresponding grasp features recognized in both images have to be detected by use of image and eventually motion information. At first, pairs of parallel contour segments with similar curvature, length and distance are searched. If a definite correspondence cannot be achieved only by these contour features, the information about the covered distance for generating the second image is used to select the grasp feature which is located closest to the expected one.

But object fine inspection requires the 3D profile of the extracted grasp features emphasized by their contour normal lines in fig. 5b. The length of these normal lines indicates the depth, the distance between a contour point and the camera. In order to determine this 3D profile, correspondent points have to be detected. Again, the information about the covered motion is used to predict the position of a corresponding point. If the correspondence is ambiguous, the relative position

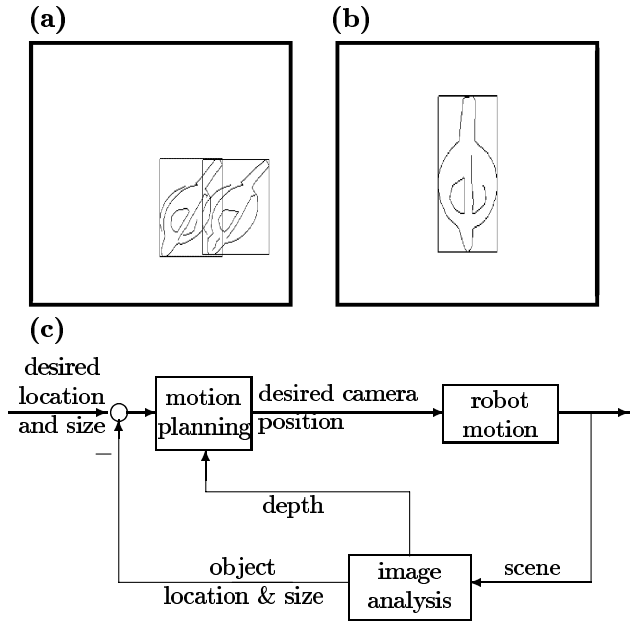

Figure 4: *Starting from an arbitrary camera position a small camera translation is performed (a) in order to align the object to the image center in an upright position (b). Inaccuracies of the vision-based motion planning are compensated in a control loop (c).*

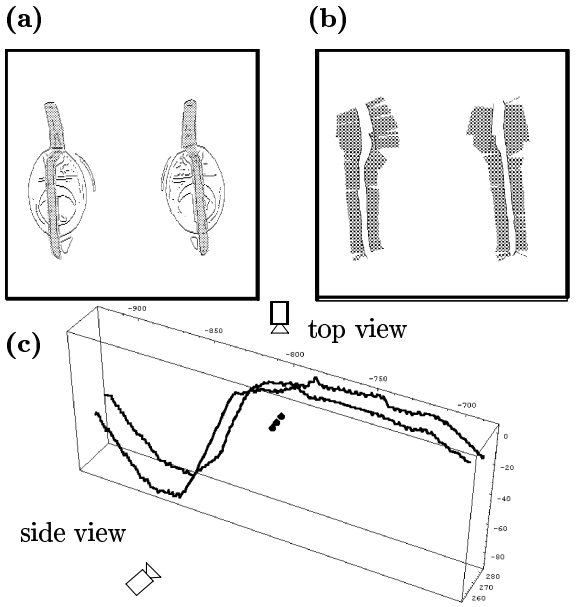

Figure 5: *Expanded grasp features shadowed in grey are extracted from stereo contour images of a watering can (a). The length of the grey coloured lines printed perpendicular to the feature contours symbolize the depth of each contour point (b). The 3D profile of the grasp feature allows to navigate the camera from a top view position to a side view position (c).*

to a curvature discontinuity in the neighbourhood of the examined point is used as an additional feature to find the correct correspondence.

3D localization of corresponding points is performed by triangulation. The 3D profile of the watering can in fig. 5c clearly shows the necessity for localizing more than one point of the selected grasp feature. Regarding the spout of the watering can, combined with the handle to one large grasp feature, the discontinuity in the 3D profile may cause problems for grasping. Using the determined 3D position and orientation of the selected grasp feature (indicated by the dotted line in fig. 5c) a new camera position can be planned allowing to inspect the side view of the object.

3.3 Combining Multiple Views

Although a 3D localization of contour parts only from the top view is sufficient for grasping many planar objects to be contacted by fingertips only, a secure grasping of volumetric objects requires an entire object inspection from additional views. But the composition of information acquired from different views is a difficult problem to be solved. For recognition of corresponding grasp features in different views an interesting approach (Taylor *et al.*, 1994) tracks antipodal grasp points for a parallel-jaw gripper during small camera motions, until the object is entirely inspected. But complex contour information describing favourable contact areas as shown in fig. 5, can hardly be tracked in real-time during camera motion. Therefore, in this approach a relatively large camera motion is performed in order to examine the object from a side view (fig. 6a) and perceive more information about the areas around selected grasp features. Guided by the grasp feature localized in the top view, which provides the largest contact area for the fingers in dexterous joint positions, the camera is navigated in front of this feature. For this, the covariance matrix of the 3D feature points is determined. The eigenvectors of this matrix describe the 3D orientation of the feature (indicated by the dotted line in fig. 5c). With respect to the 3D position and orientation of the selected grasp feature, a new camera position can be planned within which the object can be inspected from a side view.

In this side view, grasp feature extraction and localization is performed as described in sections 2 and 3.2. If the same feature, e.g. the handle of the watering can (fig. 6b), has been selected in both views, it provides sufficient accessibility for the fingers of the gripper on its top face as well as on its lateral face. This is a necessary condition for enclosing the handle by the fingers. The spout of the watering can, for example, is regarded as suitable grasp feature in the top view (fig. 5), but from the side view it is rejected and only the handle is extracted because the lateral contours of the spout are not in parallel along a sufficient distance (fig. 6).

For composition of the top view profile (full line in fig. 7) and the side view profile (dashed line in fig. 7), those feature points have to be selected describing the common part of the handle which were favoured from the top view as well as from the side view. This common part of the different 3D feature profiles is called grasp primitive. So, the grasp primitive in our example comprises only the feature points describing the handle of the watering can. In fig. 7 this grasp primitive is enclosed by a grey coloured cube symbolizing the accessible grasp space around the primitive. The fingers can be inserted into the grasp space of this primitive in order to enclose it.

In case that no correspondence between features

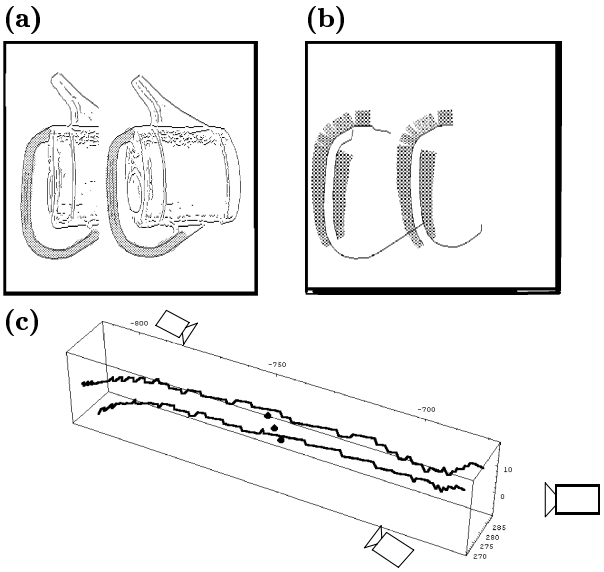

Figure 6: *Stereo contour images (a) from a side view. Expanded grasp features are emphasized by their normal lines indicating the depth (b) and its 3D side view profile can be used to plan alternative camera positions (c).*

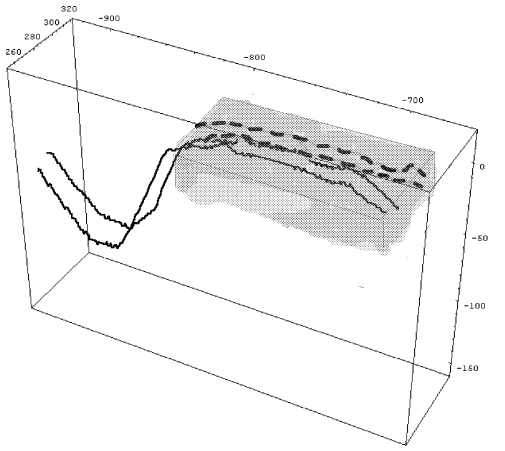

Figure 7: *The 3D profile determined from the top view (full lines) and this localized from the side view (dashed lines) are combined to a resulting grasp primitive enclosed by the grey shadowed grasp space.*

of two views could be detected or no acceptable grasp feature could be found at all, basically the same steps of feature extraction, localization and composition have to be performed again using alternative side views.

4 AUTONOMOUS OBJECT HANDLING

Beside object information acquired visually, the grasp goal to be specified by the user and constraints due to the gripper geometry are fundamental aspects for reasoning about the specific grasp situation. The information about the selected grasp primitives extracted from several views combined to a 3D representation of the favourable grasp space allows a decision, whether the gripper should approach the object perpendicular to its top or to one of its side faces. Beside the decision on the grasp plane, within which the fingers should contact the object, the grasp mode has to be determined. The employed system is able to plan precision grasps using fingertips only, as well as power grasps using finger links and/or the palm for enclosing the object, e.g. through holes, dependent on the perceived object shape.

4.1 3D Grasp Planning

Object inspection from different views ensures collision avoidance, accessibility of the lateral contact faces and the possibility to enclose the grasp primitive by the fingers simply because the primitive is visible from several views. If there was no hole below the handle at the watering can, also no parallel contour segments would have been visible in the side view and a grasp feature would not have been found. Although the object could be grasped from above using the fingertips only due to the features detected in the top view, the grasp security is inreased if the accessibility of the lateral contact faces is proved by object inspection from a side view. Furthermore, the possibility is given to approach the gripper perpendicularly to the lateral contact faces. Providing a selection of the preferred grasp plane the flexibility of the planning system is increased, also. If the major goal of the given task is to perform a very stable grasp, e.g. for object handling, power grasping is preferred. In case that the grasp should provide a maximum dexterity, e.g. for object manipulation, a fingertip grasp would be preferred.

However, the finger approach and contact phase has to be planned by determining favourable finger positions in the grasp space. For determination of the optimum contact points, the 3D points of the composed grasp primitive are examined.

Regarding the 3D contour points of the grasp primitive localized from the top view (full line in fig. 7) a set of three points is searched forming the maximum isosceles grasp triangle, which is possible due to maximum gripper opening width. The further the distance between the contact points, the larger the area which can be enclosed by the fingers. Therefore the search strategy systematically checks each isoceles triangle of supposed contact points if their distances are within the limits of the gripper work space. After examination of the contour points localized from the top view, those localized from the side view (dashed line in fig. 7) are examined as well. The contact points forming the largest isosceles grasp triangle are selected as optimum contact point configuration. If the optimum contact points were localized from the top view, the gripper approaches the primitive from above. If they were localized from the side view, the gripper approaches perpendicularly to the inspected lateral contact face. So, the gripper approach direction is always in parallel to the viewing direction.

As an example, the contour of the grasp primitive localized from the top view is illustrated in fig.8 assuming a grasp from above. Before the gripper moves down to grasp the object, the fingers are preshaped in order to avoid collisions with the grasp primitive. Preshaping positions P_1, P_2, P_3 for the fingertips are planned to be located in front of the lateral contact area (coloured in grey in fig. 8) with respect to the selected contact points C_1, C_2, C_3. From this preshaping configuration, the fingertips have to be moved to C_1, C_2, C_3 in

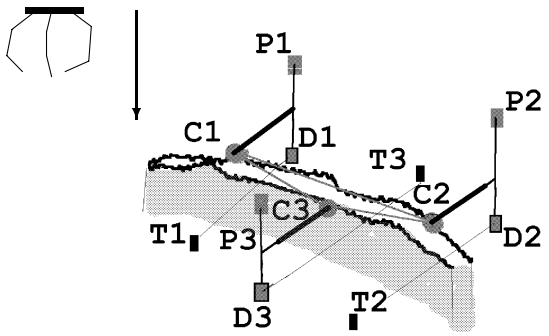

Figure 8: *Assuming a grasp from above, optimum contact points C_1, C_2, C_3 are selected and corresponding positions for preshaping P_1, P_2, P_3 are determined. The robot is moved down until a collision is sensed in positions D_1, D_2, D_3. By stiffness controlled grasping the fingertips are moved to the target positions T_1, T_2, T_3.*

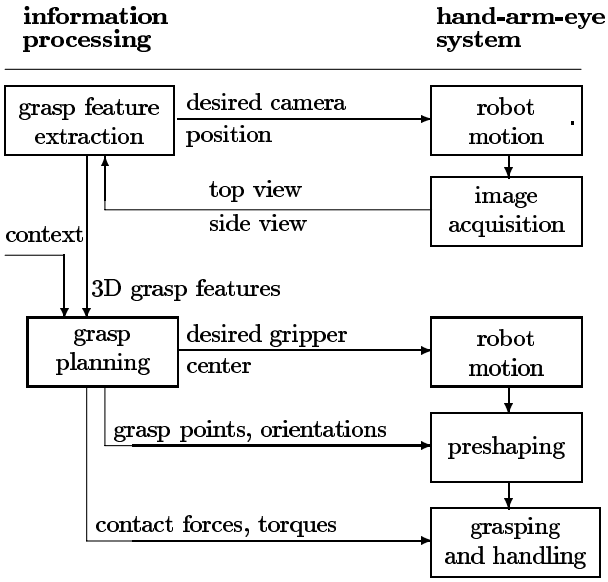

Figure 9: *Interface and system structure of the hand-arm-eye system.*

case of a fingertip grasp. Power grasping requires a more detailed planning as well as a coordinated control of hand and arm.

4.2 Hand-Arm Control

Based on the acquired object and context information, the grasp planning system has determined optimum contact points as well as a corresponding position of the gripper center (fig. 9).

In order to enclose the object, the fingertips have to be moved around the grasp feature as far as possible. Therefore, after preshaping the robot continues the approach and inserts the fingertips deeper into the grasp space. This motion is constrained either by the floor on which the object is located or by some object parts located below or behind the grasp feature. Regarding the expamle of the watering can, the gripper is moved down until the fingertips collide with the object part below the handle of the watering can. The collision is sensed by the joint torque sensors and the gripper is immediately lifted by a compliant robot motion. This coordinated hand-arm strategy allows to position the grasp feature deeply in the work space of the gripper. Assuming that the fingertips are located in positions D_1, D_2, D_3 (fig. 8) after lifting the gripper, some target points T_1, T_2, T_3 are planned as reference position for the fingertips. To achieve an enclosure of the handle by the fingertips, the target points are located on the other side of the contour than the contact points themselves (cf. fig. 8).

For the sake of a stable grasp, stiffness control is employed in order to exert contact forces and compensating torques to the object to be handled. During the stiffness controlled motion starting in points D_1, D_2, D_3 to the target points T_1, T_2, T_3, the joint torques in the fingers are supervised, until the target positions and the desired contact forces are achieved.

5 EXPERIMENTAL RESULTS

The basic result of this work is that grasping arbitrarily shaped, volumetric objects can be reliably planned using a mobile camera for acquisition of adequate object information. The hand-arm-eye system we used for inspection and the performance of those grasps is not ideal but useful to obtain qualitative experimental results.

Figures 10a-d summarize the characteristic steps of the experiment. Initially, the object of interest is zoomed to an optimum representation in the image (fig. 10a). The procedure works stably if the object is entirely visible in the initial image. Therefore, extremely large objects could not be analyzed because the distance between object and camera is constrained due to the limited robot work space.

Automatic object zooming is required for isolating the object to be grasped within the scene, because the following analysis assumes that no other object is visible. For grasp feature extraction, image processing has to be more sensitive than during zooming, because otherwise interior object edges, which may be selected as favourable grasp features, e.g. the handle of the watering can (fig. 5), would not be visible in the contour image. This results in the dilemma that the sensitivity should be low in order to suppress noise and reduce computing time, but on the other hand it should be high to detect the interior object edges. To solve this problem, a local light source is employed illuminating the object in the same direction as the camera views. This ensures that edges in the front are always brighter than the background. Nevertheless, threshold and parameter adjustment is necessary to visualize the interesting edges. However, it is sometimes hard to recognize two contour segments as a favourable feature, because a distinction between parallel and non-parallel segments has to be decided by static thresholds. So, the performed feature extraction contradicts human reasoning, in few cases. But usually, the powerful contour filtering and enhancement allows to detect every visible grasp feature, especially because most objects comprise only one contour part providing sufficient large contact areas.

In principle, the 3D profile of the features extracted from the top view (fig. 10b) allows to grasp the object approaching the gripper perpendicularly to the top view and contacting the feature

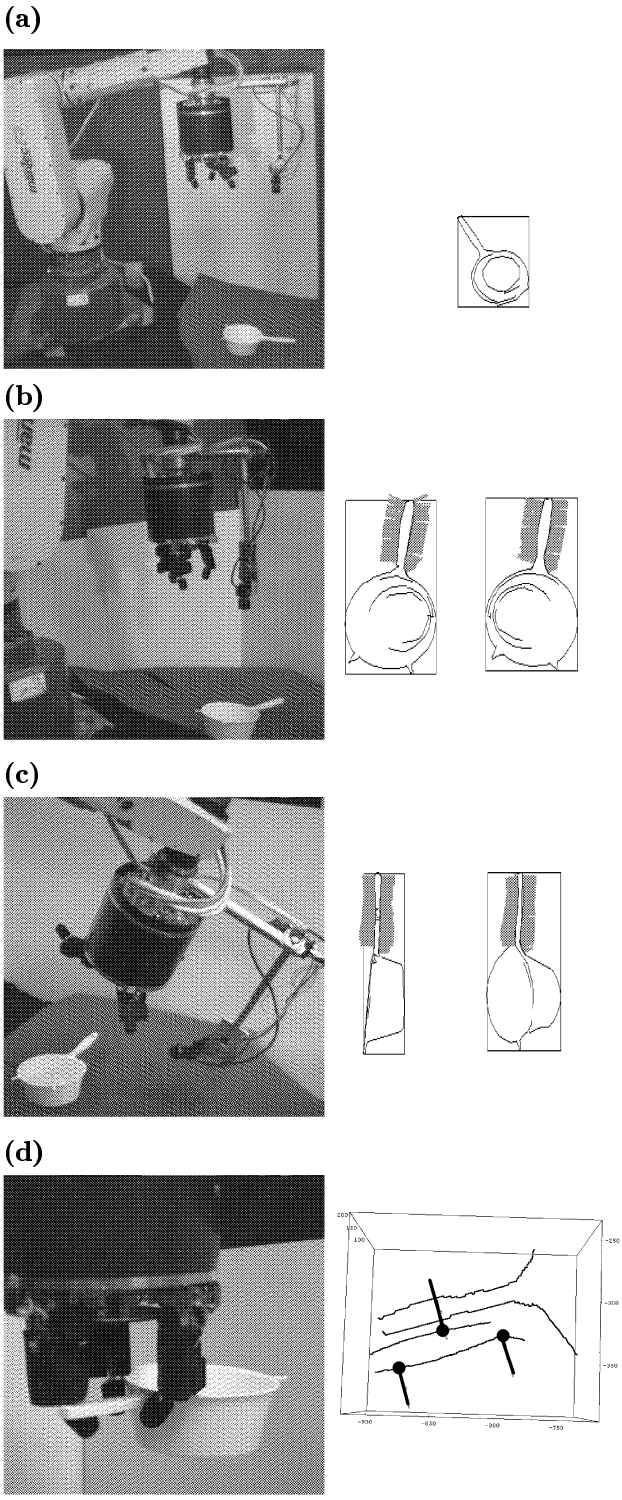

Figure 10: *Basic steps of the experiment are automatic object zooming (a), object inspection from the top view (b), object inspection from a side view (c) and object grasping due to the composed information from both views (d).*

by the fingertips, only. But for the sake of more stability, especially in case of excentric grasping, cf. e.g. fig. 10d, additional views are required for planning the enclosure of suitable object parts by the fingertips.

Due to the assumption that no disturbing contours should be visible, the background also for the side view has to be completely dark. To inspect the side faces of an object, the camera can be tilted. To avoid that the fingers of the gripper are visible in the scene during object inspection, the desired motion of the camera is divided into a desired motion of the arm and a remaining motion performed by tilting the camera (fig. 10c).

Searching favourable contact points in the 3D profile of the combined grasp primitive, the optimum configuration is determined (fig. 10d). Although grasping does not require high precision, a minimum accuracy has to be provided. The vision assisted measurements provide an average accuracy of about $1cm$. In case that the fingertips collide accidentially with parts of the object before the preshaping positions are reached, the force/torque sensor information allows to move the colliding finger away from the palm center. This ensures that the fingers are moved around the grasp feature even if the preshaping positions were inaccurately planned.

Grasping provides sometimes insufficient stability for object handling because the grasp space around the grasp feature is too small to insert the fingers deeply in order to enclose the handle of the kitchen-bolter in fig. 10d, completely. Due to the limited length of the handle and the small distance to the bottom, a classic power grasp could not be achieved but by hooking the fingertips below the handle it could be contacted by the finger links. Reasons for the limited grasp space are the thickness of the fingers and the proportions of object and gripper. Since objects like the kitchen-bolter are made for the human hand which is much smaller than the employed gripper, it is sometimes also difficult to insert the fingers into object holes, e.g. at handles (cf. fig. 1). Nevertheless, stable grasps can be performed contacting a handle by the lateral finger links (fig. 10d).

The entire size of the object area contacted by the gripper links is a coarse measure for the stability of a power grasp. Especially in case of excentric grasping, link contacts are necessary for jamming the object between the fingers and for compensating the torque due to gravity of the grasped kitchen-bolter, which may destabilize the grasp. Experiments show that the size of contact area and the stability of the grasp depends on the motion of the closing fingers. Grasp performance has not been optimized in this approach. As explained in section 4.2, the fingers are simply closed by a stiffness-controlled cartesian motion of the fingertips to a planned target position. An improvement could be achieved by an alternative contact strategy (Paetsch *et al.*, 1993) closing the finger joints separately instead of cartesian motion control as described in section 4.2.

Experiments have been performed grasping objects of every-day-life e.g. cylinders, flower pots, watering cans, kitchen-bolters, etc. Due to the limited work space of our hand-arm-eye system, there were only some locations, which allowed to move the system around the object for inspection and to approach and to grasp the object in the

favoured mode. Therefore, some camera positions, mainly for a side view, cannot be reached by the hand-arm-eye system as well as many gripper positions. So we got rare experimental results for grasping an object, if the gripper has to approach in direction of a side view. Fortunately, most objects can be also grasped, if the gripper approaches in direction of the top view, as in case of fig.10d. The expense of planning time depends on the complexity of the inspected contour and the number of selected grasp features. In most cases only one grasp feature is detected due to the shape of every-day-life objects. Although the algorithms are not optimized, the complete experiment is performed within a few minutes including the time during robot motions, which is acceptable for an offline planning system.

Apart from the specific problems due to the employed hand-arm-eye system, the experiments documented by fig. 10 show the feasability of the presented vision assisted grasp planning system.

6 CONCLUSIONS

The presented system is a proposal for an autonomous robotic system able to perform flexible grasping and object manipulations using vision for grasp and motion planning. Autonomous inspection and handling of unmodelled common objects of every-day-life is especially required for applications of service robots in the real world. Experiments documented the capabilities of the presented hand-arm-eye system in object inspection and handling and revealed practical problems which were discussed.

Most of these problems restricting a robust feasability are due to the limited robot work space and the difficult accessibility of grasp features. A service robot carrying such a hand-arm-eye system provides a sufficient accessibilty for reaching every object to be grasped and handled. Although the paper presented strategies for planning power grasps on arbitrarily shaped objects, additional strategies are required in order to improve the graspability of objects in some situations. As discussed, the size of the gripper is one reason for a restricted accessibility of grasp features. So, a coffee mug with a relatively small handle cannot be stably grasped by our system. But at least one finger can be moved into the hole of the handle in order to hook the handle. In our future work we want to focus on the stabilization of those grasps by online supervision of the contact phase after hooking.

But the graspability can also be restricted by the insufficient accessibilty of the object itself. For example, a pen lying on a table can only be grasped by a fingertip grasp. An additional manipulation is required moving the object inside the hand in order to contact the pen stably by several links of the hand. In this context we will examine the limits of a global visual sensor in comparison to local tactile sensors.

Another difficulty employing a service robot to the real world is the complexity of the scene in which objects have to be tracked. The presented algorithm for automatic object zooming can track an isolated object. In order to identify and isolate an object, colour vision may be helpful for segmentation and object isolation in a complex scene. Although the presented grasp feature detection is independent of the specific object representation, e.g. colour, shape, etc., it is required that only the object is visible in the scene. For future applications of our hand-arm-system, we will examine the advantages of colour vision, in particular for a quick separation of the object from the background of the scene.

ACKNOWLEDGEMENT

The work in this paper has been prepared in connection with the research grant To-75/18-2 by the *German National Science Foundation (DFG).*

REFERENCES

Bard, C., Laugier, C. and Milesi-Bellier, C. (1994). An integrated approach to achieve dexterous grasping from task-level specification. In: *Proc. of the Int. Conf. on Intelligent Robots, IROS'94.* Munich, Germany.

Ferrari, C. and Canny, J. (1992). Planning optimal grasps. In: *Proc. of the Int. Conf. on Intellogent Robots and Systems, IROS'92.* Raleigh, USA.

Kaneko, K. and Honkawa, K. (1994). Contact point and force sensing for inner link based grasps. In: *Proc. of the Int. Conf. on Robotics and Automation.* San Diego, USA.

Paetsch, W. and Kaneko, M. (1990). A three fingered, multijointed gripper for experimental use. In: *Proc of the IEEE Int. Workshop on Intelligent Robots and Systems (IROS'90).* Tsuchiura, Japan.

Paetsch, W., Buck, M., Weigl, A. and Tolle, H. (1993). On realising various grasps with a given dexterous robot hand. In: *Proc. of the Int. Conf. on Advanced Mechatronics.* Tokyo, Japan.

Park, Y. and Starr, G. (1992). Grasp synthesis of polygonal objects using a three-fingered robot hand. In: *Int. Journal of Robotics Research.* Vol. 11, No. 3.

Reznik, D. and Lumelski, V. (1994). Multi-finger 'hugging': A robust approach to sensor-based grasp planning. In: *Proc. of the Int. Conf. on Robotics and Automation.* San Diego, USA.

Seitz, M. and Kraft, J. (1994). Some approaches to context based grasp planning for a multi-fingered gripper. In: *Proc. of the Int. Conf. on Intelligent Robots and Systems, IROS'94.* Munich, Germany.

Taylor, M., Blake, A. and Cox, A. (1994). Visually guided grasping in 3d. In: *Proc. of the Int. Conf. on Robotics and Automation.* San Diego, USA.

Weigl, A. and Seitz, M. (1994). Vision assisted disassembly using a dexterous hand-arm system: An example and experimental results. In: *Proc. of the Symp. on Robot Control, SYROCO'94.* Capri, Italy.

Yoshimi, B. and Allen, P. (1994). Active, uncalibrated visual servoing. In: *Proc. of the Int. Conf. on Robotics and Automation.* San Diego, USA.