An Intelligent Dual Mode Vision Guided Robotic System

Kevin G. Stanley
NRC-IFCI
Vancouver
BC, Canada
Kevin.stanley@nrc.ca

Q.M. Jonathan Wu
Dept. of Electrical Engineering
University of Windsor
Ontario, Canada
jwu@uwindsor.ca

William A. Gruver
School of Engineering Science
Simon Fraser University
Burnaby, BC, Canada
gruver@cs.sfu.ca

Abstract - *Industrial robotics have looked to vision systems for flexibility. This promise has largely been unrealized because existing systems are either too slow or too inaccurate. Both visual servoing and traditional look and move are insufficient because visual servoing requires too much bandwidth, and look and move requires very accurate calibration. To mitigate these effects, we have designed a hybrid system. Our hybrid system is composed of a roughly calibrated look-and-move system using a linear approximation, and a gain scheduled PD controller which performs visual servoing. The system performs markedly better than visual servoing or look-and-move techniques in isolation. This system will have many potential applications including bin-picking, sorting, and tele-operation.*

Keywords: *Visual servoing, vision-guided robotics, robot vision, robot control.*

1 Introduction

Vision guided robotics has a rich research history that dates back to the late seventies and early eighties. While many elements of vision guided robotics have been thoroughly researched, few vision guided robotic systems have found their way into industry. Most systems were too slow and too sensitive to the environment to be useful in an industrial setting. With the rapid rise in computing power and the drop in price of high quality robotic and vision systems, the application of vision guided robotic systems to an industrial setting is becoming a reality. However, there are still barriers and limitations to the production of generic, robust, and practical vision guided robotic solutions. In this paper we propose a potential solution based on a hybrid approach.

An overview of visual servoing is given by Hutchinson, *et al*. [1] who describes the research and fundamentals of geometric feature based visual servoing. Corke [2] has shown that the performance of visual servoing algorithms can be enhanced by incorporating the dynamics of the system in the model. Papanikopolous, *et al*. [3] have used adaptive control techniques to perform visual servoing.

Calibration of the kinematic transforms between the image and the world coordinate system has been examined by many researchers. The calibration-based method has been based on determining the coefficients of the transform using captured from the robot camera pair. Wang [4] described in detail the relationships between the different frames and applied three different methods to approximate the transform, ranging from the known target and position case, to unknown target and position. Horaud, *et al*. [5] described the effect of perspective model on the accuracy of the approximation. Wei, *et al*. [6] outlined an approach for computing the transform based on active vision principles. Zhuang, *et al*. [7] described a system where both the robot and camera were calibrated simultaneously. Remy, *et al*. [8] simplified the estimation by employing Euler representations in the transform. Our approach is presented in [9] and [10].

A great deal of research efforts have been made in visual servoing recently, such as those presented in [12]-[14]. Since these do not directly relate to our work, they are not discussed in detail here.

A core problem in vision guided robotics is balancing the requirement for both accuracy and speed. Accuracy requires that many samples of the target position be taken throughout the motion of the robot. Speed constraints require that few iterations be taken before the robot reaches its target. While the cost of computers and high-frequency vision systems is rapidly decreasing it will still be some time before the cost of additional computing power is negligible with respect to the performance gains. To increase the speed of the response, it is necessary to reduce the amount of computation by reducing the number of iterations. A balance between the speed of execution and the accuracy of positioning must be obtained. We propose a system based on a hybrid computed-kinematics and visual servoing system.

Our hybrid system is composed of a roughly calibrated look-and-move system and a gain scheduled PD controller. We demonstrate that considerable performance gains can be derived using a course look-and-move motion prior to starting to servo.

2 Hybrid system

2.1 Industrial problem statement

The purpose of vision guided robotics is to position a robot end-effector at a given offset to a target that is visually constant, but spatially variable. That is, the target always appears the same, but may deviate from its assumed location. The problem becomes one of isolating the object in an image, then using that information to position the robot with respect to the target. A general spatial solution requires six-degrees of freedom (three positions and three orientations) to be resolved. We examine the simpler four-dimensional case where the target is constrained to a surface limiting orientation to roll and position in x and y. Allowing object height to vary, there are a total of four degrees of freedom.

Once the target has been isolated in the images, the robot must use the information to move the robot to the

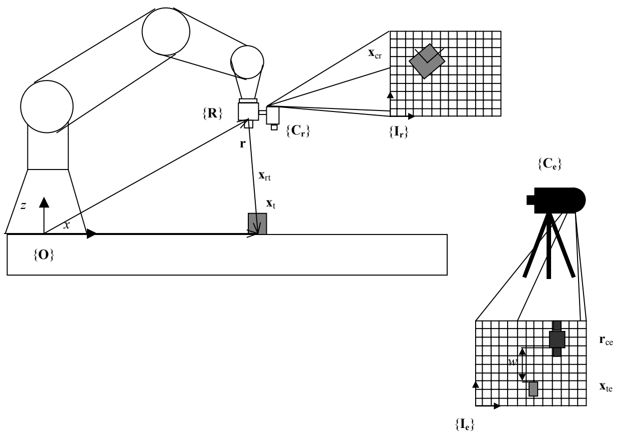

Figure 1: Frame and variable definition

desired relative position. Once there, the robot can grasp, weld, solder or rivet the target, depending on the application and the type of end-effector.

The determination of desired motion can be calculated by attempting to calculate the position of the part with respect to the robot from a single image. This is the look and move approach. While fast, it is very sensitive to calibration. Visual servoing steps the robot closer to the target, feeding back a new image after every step, resulting in a more accurate but slower response.

Vision guided robotics allows the robot to react to changes in its environmental model, such as manipulating parts arriving randomly distributed on a conveyor, or adjusting a rivet gun to match a hole in an aircraft fuselage.

2.2 Technical problem statement

The system is based on a discrete part manipulation scheme. Targets are assumed to be locally planar. The purpose of the system is to grasp a target at some unknown position $\mathbf{x}_t$ by positioning the end-effector of the robot to a position $\mathbf{r}$ with respect to the target. The position of the robot $\mathbf{r}$ is known from the encoder positions and the inverse kinematics. The target position, $\mathbf{x}_t$, must be estimated from the image systems. The x, y, and θ components of the target position, $\mathbf{x}_t$ are estimated by measuring the position of the target in $\mathbf{x}_{cr}$ in the end-effector mounted camera frame $\{\mathbf{I}_r\}$. Because targets may have varying heights, we must also estimate the distance along z in the world coordinate frame $\{\mathbf{O}\}$. This measurement is made by examining the image frame of the external camera $\{\mathbf{I}_e\}$. The difference in y in $\{\mathbf{I}_e\}$ between the observed robot position $\mathbf{r}_{ce}$ and the measured target position $\mathbf{x}_{te}$, generates w, the observed difference in world z. A diagram of the system is shown in Figure 1.

Combining $\mathbf{x}_{cr}$ with w as measured in $\{\mathbf{I}_e\}$ we can generate image space position of the target, denoted $\mathbf{x}_{ti}$, as:

$$\mathbf{x}_{ti} = (u\, v\, w\theta)^T \tag{1}$$

The problem then resolves to using $\mathbf{x}_{ti}$ to position the robot with respect to the target. This motion can be generated using either a computed kinematics (also called look-and-move) or a visual servoing. In computed kinematics, the position of the target in the world frame, $\mathbf{x}_t$ is estimated using a transform $\mathbf{T}$ such that:

$$\mathbf{x}_t = \mathbf{T}\mathbf{x}_{ti} \tag{2}$$

Determining the transform T is the key focus of [0]-[0]. Because look and move techniques are so sensitive to the calibration of $\mathbf{T}$, we also use a visual servoing system. In visual servoing the image position of the target, $\mathbf{x}_{ti}$, is regulated to a desired position $\mathbf{x}_{tid}$, such that the three-dimensional position of the target relative to the robot in the base frame $\mathbf{x}_{rt}$ is known.

$$\mathbf{x}_{rt} = \mathbf{r} - \mathbf{x}_t \tag{3}$$

That is, if

$$\mathbf{x}_{ti} = \mathbf{x}_{tid} \tag{4}$$

then Equation (3) must hold.

We have assumed that the parts will be small, discrete, rigid, and stationary. Objects commonly found in industrial or office environments are suitable targets.

2.3 Look and move

In the look and move class of vision guided robotic systems, the pose of the target is calculated from the image, and the robot is commanded to move to some point in space with respect to the target. In order to calculate the pose accurately, the kinematics between the camera and the robot must be known with a high degree of precision, and the camera parameters must also be known with a high degree of accuracy. Several adaptive self-tuning schemes have been suggested [4]-[8].

It should be apparent from Figure 2 that linear calibration is not accurate enough to generate a reliable position with respect to the target. There are two sources for this error. First, *x* and *y* vary inversely with *z*, so errors increase as *z* decreases. Second, the sensitivity to pitch and yaw errors increases closer to the target. Instead of a simple linear transform, the system is governed by:

$$\mathbf{x}_{\mathrm{t}} = \mathbf{T}\mathbf{x}_{\mathrm{ti}} \qquad (5)$$

where **T** is the transformation matrix between the image and robot frames. The only method for resolving these issues within a look and move framework is much more accurate calibration. Now all six variables of the calibration matrix must be derived. However, these variables are coupled nonlinearly, increasing the complexity and sensitivity of the calibration procedure [4]. Instead of more accurate calibration we use visual servoing.

2.4 Visual servoing

Visual servoing encompasses all the approaches for vision guided robotics where a control law is used to position a robot with respect to a target. Control laws range from simple proportional control laws to very complex adaptive schemes [1]-[3]. The control law can be phrased in terms of the image, or in terms of a position in space. In the first circumstance, image based visual servoing; the control law directly minimizes the error between the current image and the desired image. In the second, Cartesian visual servoing, the pose of the target is calculated and the relative pose between the robot and the target is regulated. The control law can also be directly run on the robot or run with the original robot controller still in the loop. Our visual servoing system is an image based, PD controller run with the robot joint controllers in the loop.

The most straightforward method for doing so is outlined in Hutchinson, *et al* [1]. Defining the motion of the camera as

$$\dot{\mathbf{C}} = \Omega \times \mathbf{P} - \mathbf{X} \qquad (8)$$

where **C** is the speed of the camera, **Ω** is the rotational velocity of the end-effector, **P** is the position of the camera with respect to the robot end-effector and **X** is the translation velocity of the end-effector. Given that we have constrained servoing to only 4 DOF, ω_x and ω_y are zero therefore:

$$\begin{bmatrix} \dot{u} \\ \dot{u} \\ \dot{w} \\ \dot{\theta} \end{bmatrix}_{image} = \begin{bmatrix} \frac{c_x}{z} & 0 & -\frac{u}{z} & -v \\ 0 & \frac{c_y}{z} & -\frac{v}{z} & u \\ 0 & 0 & c_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \omega_z \end{bmatrix} \qquad (9)$$

where *u* is the *x* coordinate of the image frame, *v* is the *y* coordinate of the image frame, and *w* is the *z* coordinate of the external image frame. The inverse of this upper triangular matrix gives the inverse Jacobian that is the plant model for this controller.

2.5 Hybrid system integration

To achieve a rapid response, and a stable solution, we have employed gain scheduling for our PD controller. The gain scheduling system uses two gain levels, a high gain approach, and a low gain positioning, to achieve a fast and accurate final position. A good example of the response of the system is shown in Figure 5.

Our visual servoing system is an image based, PD controller run with the robot joint controllers in the loop.

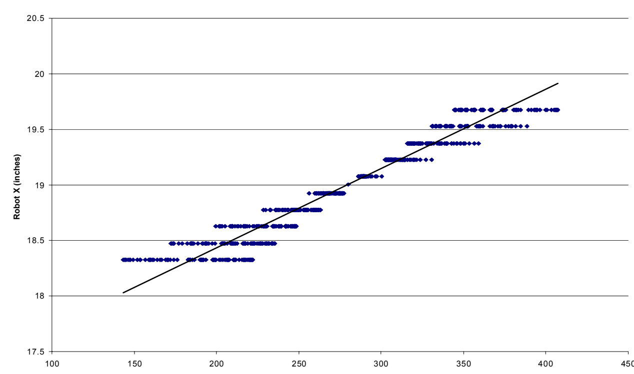

Figure 2: Measured vs predicted position

The control system is based on a classic PD controller structure as shown in Figure 3,

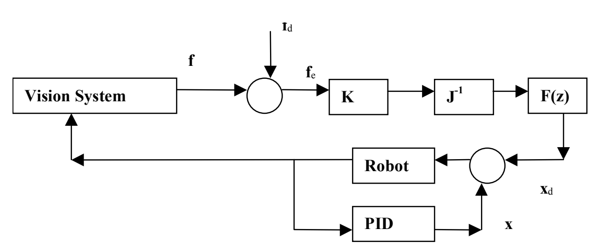

Figure 3: PD control system

where x_d is the desired position x is the measured position, f is the current feature vector, f_d is the desired feature vector, f_e is the current feature error, **K** is the gain matrix, $\mathbf{J}^{-1}$ is the inverse feature Jacobian and $F(z)$ is the discretization function. The plant model in this case is the inverse feature Jacobian. It is a mapping from the input feature space to the output Cartesian space. The Jacobian maps image errors to Cartesian velocities. The Jacobian is derived from the geometry of the system. Our derivation of the image Jacobian is closely related to Hutchinson's [1].

Defining the motion of the camera as

$$\dot{\mathbf{C}} = \Omega \times \mathbf{P} + \mathbf{T} \quad (1)$$

Where **C** is the speed of the camera, **Ω** is the rotational velocity of the end-effector, **P** is the position of the camera with respect to the robot end-effector and **T** is the translation velocity of the end-effector. Given that we have constrained servoing to only four DOF, ω_x and ω_y are zero therefore:

The inverse feature Jacobian is:

$$\mathbf{J}^{-1} = \begin{bmatrix} \frac{w}{c_x c_z} & 0 & \frac{u}{c_x c_z} & \frac{wv}{c_x c_z} \\ 0 & \frac{w}{c_y c_z} & \frac{v}{c_y c_z} & -\frac{wu}{c_y c_z} \\ 0 & 0 & \frac{1}{c_z} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

The measurement of z in the inverse Jacobian is stated in terms of the image measurement of z (w) because the distance from the manipulator to the top of the target is not known.

2.6 Gain scheduling

Control strategies generally must balance the precision of convergence with the number of iterations required to converge. While higher gains tend to converge faster, they can lead to oscillation and even limit cycling. Because each iteration is computationally expensive, we must reduce the number of iterations to converge to a solution in a timely fashion. To achieve a rapid response, and a stable solution, we have employed gain scheduling for our PD controller.

The gain scheduling system uses two gain levels, a high gain approach, and a low gain positioning, to achieve a fast and accurate final position.

The lower gain system removed the limit cycling behavior, which primarily occurred in x and y, and occasionally caused a coupled effect on θ. By halving the proportional and derivative gains for the controller in x and y, we were able to eliminate limit cycling. The high gain system is executed after the initial computed kinematics move. The high gain system continues until the target is within a 15-pixel square of the center of the image. The high gain matrix is then swapped with the low gain matrix. The lower gain matrix has an overdamped response, which drives the robot to its final position smoothly.

As is apparent from [11], there is a marked difference in system response for the three stages of motion. The look and move portion is a linear point to point motion, the high gain servo is characterized by a rapid response and oscillatory underdamped behavior. The low gain servo is more typically overdamped and moves the robot smoothly to its final position. The three stages of motion are very apparent from studying the y response.

3 Experiments and results

The hybrid positioning algorithm is tested against a visual servoing control to evaluate the performance gain of the hybrid system against pure visual servoing. The first experiment determines the number of iterations and the final accuracy achieved for a circular target for both motion algorithms. The second experiment uses several different objects shown in Figure 4 to demonstrate the robustness of

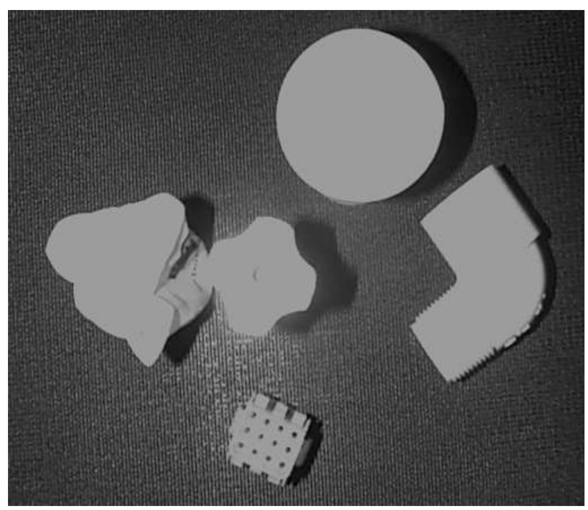

Figure 4: Targets used in experiments

the system to different target types.

The system is composed of a dual 200 MHz Pentium Pro computer with a PCI bus. It contains a Matrox Genesis frame-grabber and digital image processing board. It is connected to the CRS C-500 controller running the RAPL-3 operating system and communicating with the PC over a serial cable. All image processing is carried out on the

Genesis board. The genesis board allows multiple threads of processes to be queued via the PCI bus.

In this experiment, the robot was presented with five targets as shown in Figure 4. The object, a dixie cup, appears as a circle in the top projection, and was run eight times for each type of robotic positioning system as a baseline for observing the other targets. The dixie cup target was used to test the accuracy of the system. Of the eight target positions, each system was run until an accuracy of ± 5 pixels on each joint was achieved. The final error and number of iterations required are shown in Table 1.

Table 1: Comparison of Systems

Algorithm	Final Error	Iterations
Visual Servo	2.98	9.86
Look and Move Hybrid	2.73	6.5

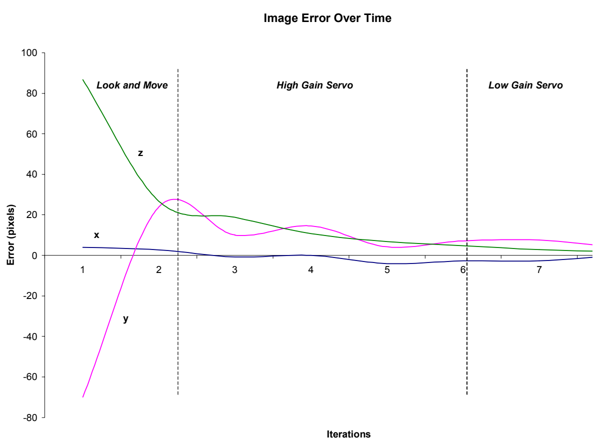

Figure 5. Sample muti-phase response for a hybrid system

It should be apparent that the visual servoing systems have approximately the same final error as the hybrid system. However, the look and move hybrid systems require 33% fewer iterations to converge. This leads to a time saving of over a second for our low bandwidth system.

The algorithms' generality with respect to target type was tested by placing different targets in the workspace, and running the algorithms. The results are shown in Table 2.

As is apparent from [11], there is little difference in accuracy (less than one pixel) between the two systems, even with different targets. There is little difference between the average values for the cup, and the measured values for the other targets. However, the performance advantage of the hybrid systems over the visual servoing system is again significant.

Table 2: Target generalization comparison

	Visual Servoing		Look and Move Hybrid	
	Iterations	Error	Iterations	Error
backing	11	0.79	7	2.48
crushed	12	2.93	12	2.67
elbow	9	3.36	5	3.06
plug	10	2.95	5	3.79
Average	10.5	2.51	7.25	3

4 Industrial Applications

Vision guided robotics are applicable to manufacturing problems where ever there is a potential for a change or variation in the position of the robot's target. Image information can be used to localize the target. Our method provides a fast and accurate method of approaching the target. Specific industrial applications include:

4.1 Bin picking

The bin of parts problem is a classic vision guided robotics problem. By enabling the robot to remove parts from a bin, costly and complicated part feeders can be eliminated. This results in capital savings and reduced downtime because the part feeders often jam. However, efficient vision guided robotics is only one part of the solution. Significant image processing must be performed to properly isolate the individual parts within the bin. This remains a significant challenge in machine vision research.

4.2 Sorting

Robots are increasingly employed in palletizing and material handling operations. By adding a vision classification system, the robot can be employed as a sorter. Articulated robots are not suitable for all sorting tasks, but those tasks requiring manipulation of the sorted item, such as palletizing, are ideally suited to vision guided robotics. Visual feedback allows the parts to arrive on a standard conveyor at varying rates and positions.

4.3 Assembly

Poorly formed parts with low dimensional tolerance, such as some injection modeled plastics, can be assembled using robotic vision. Important features on the target can be identified and used to guide the assembly process. Coupled with effective force feedback, vision guided robotics could also be used in assembly tasks such as furniture manufacturing where the precise dimensions and material quality cannot be completely controlled.

4.4 Tele-operation

Tele-operation is often plagued by slow responses and unintuitive robotic commands. Speed and ease of use could be increased if more computation could be performed at the robot site. The user could select targets on a screen, and allow the robot to grasp them without have to use awkward joystick controls. This would also reduce bandwidth requirements by removing the need for real-time video connections.

5 Conclusions

The system manipulates a wide range of discrete parts in four degrees of freedom without target models. We have demonstrated that an efficient approach to vision guided robotics with limited bandwidth is a hybrid computed kinematics and visual servoing system. Limited bandwidth applications will be the only applications in industry for several years because robotic vendors are unwilling to provide complete access to their controllers, and low cost vision systems will not have the required ability to operate at speeds on the order of 1 kHz. The technology and hardware cost now allow vision-guided robotics in practical applications.

References

[1] Seth Hutchinson, Gregory Hager, and Peter I. Corke, "A Tutorial on Visual Servo Control", *IEEE Trans. on Robotics and Automation*, Oct 1996, pp. 651-669.

[2] Peter I. Corke, Malcolm C. Good, "Dynamic Effects in Visual Closed Loop Systems," *IEEE Trans. on Robotics and Automation*, Oct 1996, pp. 671-683.

[3] Nikolaos P. Papanikolopoulos, Pradeep K. Khosla, "Adaptive Robotic Visual Tracking: Theory and Experiments," *IEEE Trans on Automatic Control,* March 1993, pp. 429-445.

[4] Ching-Cheng Wang, "Extrinsic Calibration of a Vision Sensor Mounted on a Robot," *IEEE Trans. on Robotics and Automation,* Vol. 8, No. 2, April 1992, pp. 161-175.

[5] Radu Horaud, Fadi Dornaika, Bart Lamiroy, and Stephane Christy, "Object Pose: The Link between Weak Perspective Paraperspective and Full Perspective," *International Journal of Computer Vision* 22(2), 1997, pp. 173-189.

[6] Guo-Qing Wei, Klaus Arbter, and Gerd Hirzinger, "Active Self-Calibration of Robotic Eyes and Hand-Eye Relationships with Model Identification," *IEEE Trans. on Robotics and Automation,* Vol. 14, No. 1, February, 1998, pp.158-166.

[7] Hanqi Zhuang, Kuanchih Wang, Zvi S. Roth, "Simultaneous Calibration of a Robot and a Hand-Mounted Camera," *IEEE Trans. on Robotics and Automation,* Vol. 11, No. 5, 1995, pp. 649-660.

[8] S. Remy, M. Dhome, J. M. Lavest, N. Daucher, "Hand-eye calibration," *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1997 pp. 1057-1065.

[9] K. G. Stanley, J. Wu, W. A. Gruver, "An intelligent vision guided telerobotic system for file manipulation and office automation," *IEEE International Conference on Systems, Man, and Cybernetics, Oct. 2000*, pp. 1448-1453.

[10] K. G. Stanley, J. Wu, W. A. Gruver, "A hybrid neural network based vision-guided robotic system," *IFSA/NAFIPS International Conference*, July 2001.

[11] Kevin Stanley, Q.M.J. Wu, Ali Jerbi and William A. Gruver "A fast two-dimensional image based grasp planner" *IEEE International Conference on Intelligent Robots and Systems, 1999,* Kyongju, Korea, Oct. 1999.

[12] M. Iwatsuki and N. Okiyama, "A new formulation of visual servoing based on cylindrical coordinate system," *IEEE Transactions on Robotics and Automation*, Vol. 21, No. 2, pp. 266 – 273, 2005.

[13] A. Comport, M. Pressigout, E. Marchand, and F. Chaumette, "A visual servoing control law that is robust to image outliers," *Procs. of Intelligent Robots and Systems*, Vol. 1, pp. 492-497, Oct. 2003.

[14] E. Malis and P. Rives, "Robustness of image-based visual servoing with respect to depth distribution errors," *Procs. of IEEE International Conference on Robotics and Automation*, Vol. 1, pp. 1056-1061, Sept. 2003.