A General Learning Approach to Multisensor Based Control using Statistic Indices *

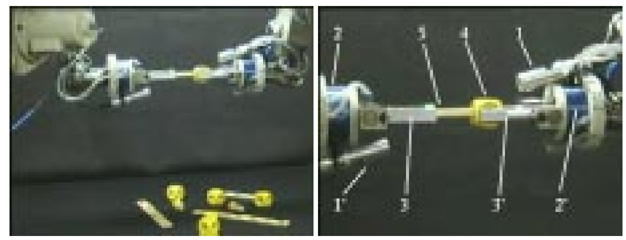
Yorck von Collani, Markus Ferch, Jianwei Zhang, Alois Knoll
Technical Computer Science, Faculty of Technology,
University of Bielefeld, 33501 Bielefeld, Germany

Abstract

We propose a concept for integrating multiple sensors in real-time robot control. To increase the controller robustness under diverse uncertainties, the robot systematically generates series of sensor data (as robot state) while memorising the corresponding motion parameters. From the collection of (multi-) sensor trajectories, statistical indices like principal components for each sensor type can be extracted. If the sensor data are preselected as output relevant, these principal components can be used very efficiently to approximately represent the original perception scenarios. After this dimension reduction procedure, a non-linear fuzzy controller, e.g. a B-spline type, can be trained to map the subspace projection into the robot control parameters. We apply the approach to a real robot system with two arms and multiple vision and force/torque sensors. These external sensors are used simultaneously to control the robot arm performing insertion and screwing operations. The successful experiments show that the robustness as well as the precision of robot control can be enhanced by integrating multiple additional sensors using this concept.

1 Introduction

In our research work on sensor-based robot control [6], we are faced with many high-dimensional problems concerning a large number of input variables which importance and inter-dependence are not clearly known. It is well-known that general fuzzy rule descriptions of systems with a large number of input variables suffer from the problem of the "curse of dimensionality". But in many real-world applications it is difficult to identify the decisive input parameters and thus to reduce the number of input variables to the minimum. Hence a general solution to building fuzzy models is not only interesting from a theoretical point, it may also extend the range of applications of fuzzy control to more complex intelligent control problems.



(a) (b)

Figure 1: The experimental setup for assembly. 1,1': hand-camera; 2,2': force/torque sensor; 3,3': parallel jaw-gripper; 4: nut; 5: screw.

1.1 Vision/Force- Guided Robot Motion

Assembly skills like inserting and screwing are part of the most important and most demanding sensor-based manipulation skills of cooperating robots. The use of force-feedback is the mostly used sensor information source in robotics but in recent years visual feedback and especially the integration of both had been of great interest. Conventional techniques try to exploit a common representation space to achieve a fused model of the environment ([7], [3]). In [11] this is achieved by describing sensor observations in terms of uncertain geometry using probabilistic fusion methods. In [1] vision together with an internal strain gauge is used to gather information about the contact forces acting on a hand during grasping. In [10] force and vision feedback is combined using so called *vision and force resolvabilities*. Another approach is presented in [2]. There the force and vision information is fused by using a task frame formalism. As an example a vision algorithm reconstructs the 3D position of a feature point, using also the distance information from a force sensor. Common to nearly all these approaches is an explicit modeling of the sensor properties in order to combine the information. In our work

*The work described in this paper is partly funded by the Deutschen Forschungsgemeinschaft in the project SFB 360/D5

presented here we fuse visual information from two uncalibrated cameras resp. from one camera and a force/torque sensor. We do not use any explicit models but employ an adaptive neuro-fuzzy scheme to *learn* the appropriate robot motions necessary to perform a complex screw-task.

A special case of vision-action transformation is camera-supported fine-motion control. *Affine Visual Servoing* [4] may be applied to such tasks. The changes in the shape of image contours are the input of a feed-forward controller. In recent years the idea of using uncalibrated cameras for visual guidance has found increasing interest (c.f.[14, 5]). However, nearly all require some initial and subsequent to perform well.

CMAC neural networks may tackle the problem of dimensionality; in [8] 12 inputs represent four joint positions of the robot, four image parameters and their desired changes. The outputs are the control signals for the four robot joints.

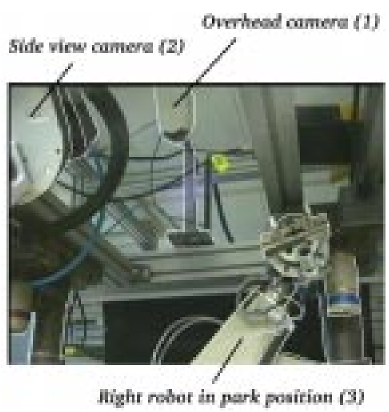


Figure 2: Location of the cameras: (1) overhead, (2) side-view, (3) robot in park position

In [9] learning of vision-based positioning based on visual appearance information was introduced. The image data set is compressed using *principal component analysis* to obtain a low-dimensional input space. A parametric eigenspace representation is used for describing the different objects as well as object locations. The positioning problem is thus transformed into finding the minimum distance between a point and a manifold in the eigenspace.

As far as we know no work on mapping the multiple images direct into "action values" has been reported.

2 Problem Description

2.1 Experiment Setup

The problem scenario (see Fig. 1 and 2), screwing a screw (5) into a nut (4) with two cooperating robots, originates from our collaborative project which aims at assembly of aggregates with wooden toy construction sets. The manipulators are installed upside down and can grasp the required assembly components from the assembly table. Each robot is equipped with a force sensor (2,2') on which a pneumatic parallel-jaw gripper (3,3') is mounted. A small camera (1,1') is fixed over the gripper. The manipulators are two Puma 260 and the host computer is a Sun UltraSPARC. We consider general screwing without using any fixture devices.

2.2 Uncertainties

For a general-purpose arm/gripper system, the following two types of uncertainties must be taken into account:

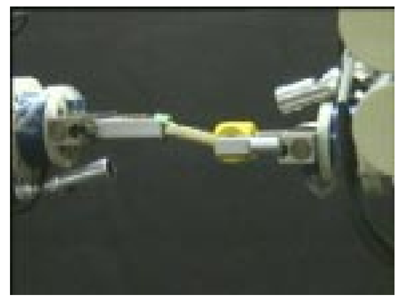


Figure 3: An inconvenient start-situation for screwing

Grasping precision. Although we have applied a hand-camera in a "self-viewing" configuration, which significantly improved the grasping precision in comparison with the open-loop positioning, regrasping still engenders deviation of the screw from the rotation axis of the gripper.

Slippage of the part in the hand. Due to the effect of the resulting forces, the screw grasped by a jaw gripper may easily slip during the screwing process.

The uncertainties in a screwing process cause the following two concrete problems:

1. The screw is not centrically grasped: the rotating axis of the screw does not match the axis of the gripper.
2. The screw is obliquely grasped, see Fig. 3.

3 Vision-Based Control

Without using sensors a screwing operation can fail under each of the uncertainties discussed above. Therefore, sensor-based compensation motions become necessary. The resulting forces in case 1. in the normal and orientation directions should be minimised and stable. Additionally, to guarantee a successful screwing-in phase, a constant force in the approach direction should be exerted. Unlike the first case, the forces and/or torques give no sufficient information about the orientation of the screw. A supplementary approach is to monitor the scene with external cameras and correct the orientation before contact is made between the screw and the nut [12]. Now we have performed two different approaches to adjust the orientation of the screw after contact is made:

1. We fuse the images from two different cameras to determine the orientation of the screw (see Fig. 5).
2. We fuse the information of a force/torque sensor and the related camera.

In both tasks we use the same B-spline neuro-fuzzy model. The displacement of the screw in the gripper is the output of the controller and hence the output can direct be used to correct the manipulator.

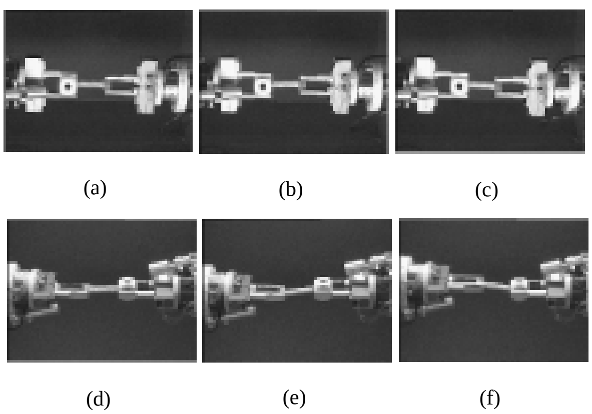


Figure 4: Typical images taken by the external cameras ((a)-(c) viewpoint from above, (d)-(f) side view

Fig. 4 shows a sequence of typical views of the scene. We therefore employed a method that extracts automatically the needed features from one or two fused images to compensate the uncertainties.

4 B-Spline Neuro-Fuzzy Model

4.1 Basic Principle

The controller for force control can be efficiently realized using the B-spline fuzzy controllers proposed in our earlier work [15, 16]. This type of controller may be characterised by the following features distinguishing it from standard fuzzy controllers:

- B-spline basis functions are employed for specifying the linguistic terms (labels) of the input variables. By choosing the order n of the basis functions, the output is C^{n-2} continuous.
- Each controller output is defined by a set of fuzzy singletons (control vertices). The number of control vertices is equal to the number of the rules and their optimal values can be iteratively found through learning. This adaptation procedure is equivalent to weight adjustment in an Associated Memory Neural Network.
- One problem with learning in conventional fuzzy controllers is that too many parameters must be adjusted. With B-spline fuzzy controllers, a simple modification of control vertices causes the change of the control surface. As far as concerned supervised learning, if the square error is selected as the quality measure, the partial differential with respect to each control vertex is a convex function. As for unsupervised learning, if the error of the cost function is approximately piecewise proportional to the error of the control values, the learning-process descent will also show stable asymptotic behaviour [15].

4.2 Dimension Reduction

It is one of our long-term research goals to find a general model which transforms raw image data directly into "action values". Our grey-scale images have 101×41 pixels and if no image processing is performed then a control system with about $4,000$ input variables (i.e. one for each pixel) needs to be modeled; the system output would be the motion values for the robot(s).

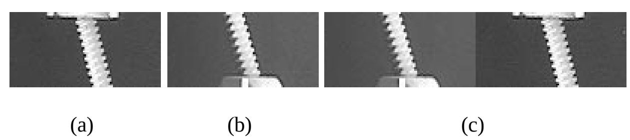


Figure 5: Clipped images from camera 1 (a) and camera 2 (b) and the resulting merged image (c).

If the dimension of the input space is small enough, the input variables can be directly covered by fuzzy sets. Each item of the rule is human readable and may be interpreted as describing a special instance of a general situation. If, however, the image of a camera is regarded as a vector, this high-dimensional sensor image is too large to build a corresponding rule base. Fortunately, sensor images are often observed in a local context: the complete situation is not of particular interest and a subspace containing all necessary information for determining the action values can be found.

4.3 Projection into Eigenspace

A well-known technique for dealing with multivariate problems in statistics is the *principal component analysis* (PCA). As shown in [9], this technique is also suitable for reducing the dimension of the input space of a general control problem. It was introduced for the use of visual learning by [13].

In the first task our approach of dealing with 3D uncertainties is to merge small and local parts of different greyscale camera images and project the resulting image into an eigenspace (see Fig. 7). In the second task we project the images from one camera into an eigenspace and take this and the data from a force/torque sensor as input for our controller (see Fig. 8). Fig. 6 shows the visualised transform matrix of the fused image date. The brighter the pixel the more relevant the component in the image.

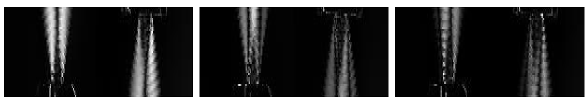


Figure 6: Visualisation of the transformation matrix: first to third principal component.

An eigenvector, denoted as EV_i, is computed as $[a_{1,i}, a_{2,i}, \ldots, a_{m,i}]^T$. The eigenvectors form an orthogonal basis for representing the original individual sensor patterns. Assume that the eigenvectors $EV_1, EV_2, \ldots$ are sorted according to their eigenvalues in a descending order. An eigenspace with a reduced dimension n can be formed with the first n eigenvectors. EV_i defines the ith dimension in the eigenspace. The projection of an input vector $X = [x_1, x_2, \cdots, x_m]^T$ onto eigenvector EV_i, called the ith principal component, is $p_i = a_{1,i}x_1 + a_{2,i}x_2 + \cdots + a_{m,i}x_m$. The complete projection can be represented as:
$[EV_1, \ldots, EV_n]^T \cdot X = [p_1, \ldots, p_n]^T$.

All projections of the sample data sequence form a manifold in the eigenspace. Such a projection can be viewed as a layer of neural network, see the connection layer of the two left parts of Fig. 7.

5 Implementation

5.1 Sampling Training Data

For training, the input data and desired output values have to be recorded. It is desirable that all typical input data be generated. As outlined above there are different orientations of the screw. For recording, the robot moves

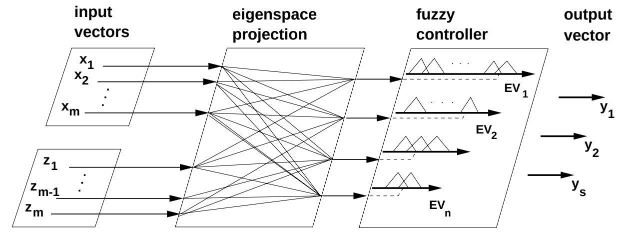


Figure 7: The structure of a fuzzy controller based on eigenspace projection by fusing images.

to the ideal position and orientation for screwing. Subsequently, it moves to several other orientations. For each of them the deviations from the ideal orientation , the forces and torques are recorded.

5.2 Calculating Eigenvectors

After the input data are sampled, the following steps are necessary:

1. The (potentially merged) images are normalised so that the energy of each image becomes 1. As an option the average image can be subtracted.
2. The images are stacked into vectors.
3. The covariance matrix of the vectors is calculated.
4. The eigenvectors and eigenvalues are calculated.
5. Each image is projected into the eigenspace.

5.3 Training the Fuzzy Controller

For the B-spline controller the training procedure is as follows:

1. Select the n eigenvectors with the largest n eigenvalues denoted as $\mathrm{EV}_1, \ldots, \mathrm{EV}_n$.
2. For the second task select the components from the force/torque vector which have the greatest relevance to the desired controller output. In this application these are the forces in N- and O- direction and torques around the N- and O- vector of the robot tool, because the rotation of the gripper is around the approach vector.
3. Select the order of the B-spline basis function for each input
4. Determine the knots of the B-spline basis functions for partitioning each input.
5. Project images onto the selected eigenvectors.
6. Initialise the control vertices for the output.
7. Learn the control vertices with the projected values from the images and the data from the force/torque sensor using the gradient descent method.

8. If the results are satisfying, terminate.
9. Modify the knots for eigenvectors, go to 5.

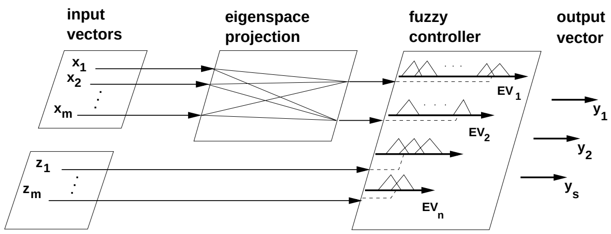


Figure 8: The structure of a fuzzy controller for fusing force and vision.

It is important to determine the right parameters for the fuzzy set. If too few eigenvectors are used, then the fuzzy controller cannot distinguish all situations. If too many eigenvectors are used, then the memory requirements of the fuzzy set and the number of required training samples is not manageable. Similarly important is the correct partition. If the partitioning is too fine, the fuzzy controller generalises insufficiently.

6 Numerical Results

The vision controller is learned with 363 training images, shifting the screw between $\pm 15^\circ$ around the N- and O-direction in steps of 3°. The learned controller is tested with additional 363 images. Fig. 9 shows the sorted eigenvalues of the covariance-matrix and it can be seen that most of the information of the images is contained in the first dimension of the reduced eigenspace.

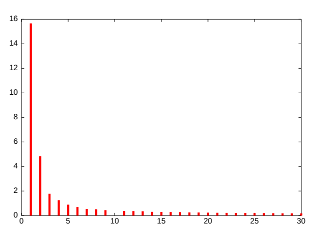


Figure 9: Sorted eigenvalues of covariance-matrix.

6.1 Fusing two cameras

After merging the two images and projecting them into the reduced eigenspace, we used the three greatest eigenvectors as input for the B-Spline fuzzy controller. Each eigenvector is covered with 10 B-Splines as membership functions.

In comparison table 1 shows the mean square error of the controller, the maximum error and the largest worst

	Mean square error $[\circ^2]$	Max. error $[\circ]$	Worst case error $[\circ]$
N- direction			
Overhead camera	8.62	7.35	7.07
Side-view camera	54.15	26.9	26.9
Fused images	2.92	6.06	–
O- direction			
Overhead camera	51.88	18.51	18.51
Side-view camera	10.0	11.33	11.53
Fused images	2.75	5.68	–

Table 1: Mean square error, maximum error and worst case error for angle around N-and O-direction.

case errors of the test images around the N- and O-direction with and without fused images. The worst case is defined as a movement of the manipulator in the wrong direction, which causes a bigger displacement.

Our first experiment shows that merging the images produces a much better controller than the separated controller for each camera and direction. The mean square error and maximum error are smaller and with the merged images in contrast there is no output of the controller, which moves the manipulator in the wrong direction. Combining this approach with the first procedure described in [16] results in a very robust and rapid technique to correct the orientation of the screw.

6.2 Combining visual and force/torque information

Simular to the first task we also use the three greatest eigenvectors from a single image as input for the B-Spline fuzzy controller. Additional the torque around the N- resp. around the O- axis of the tool is used as input (see Fig. 8). Each eigenvector is covered with 10 B-Splines and the torque with 5 B-Splines as membership functions. Table 2 also shows the mean square error of the controller, the maximum error and the largest worst case errors of the test images around the N- and O-direction with and without additional force. This experiment shows that the fusion of vision and force/torque data produce better results in comparison with the unfused case. It also shows that the B-spline neuro fuzzy model is capable of fusing different sensor data.

7 Conclusions

We have shown that the B-spline model may be utilised for sensor fusion and high-dimensional problems such as visually guided fine-motion. We have implemented the

	Mean square error $[\circ^2]$	Max. error $[\circ]$	Worst case error $[\circ]$
N- direction			
Single camera	8.62	7.35	7.07
Camera + torque	6.6	8.75	–
O- direction			
Single camera	12.9	11.64	7.65
Camera + torque	8.18	8.2	–

Table 2: Mean square error, maximum error and worst case error for angle around N- and O-direction combining force and vision.

approach with a two-arm robot system and both kinds of training are used to build the controllers: unsupervised on-line for the force controller and supervised off-line learning for the vision system.

The advantages of our approach are:

- Projecting the high-dimensional input space into a reduced eigenspace the most significant information for control is maintained. A limited number of transformed inputs can be partitioned with the B-spline model.
- By merging the different kinds of sensor data a sufficient precision can be obtained for determining the robots orientation correction.
- To solve this problem the statistical indices provide a suitable solution to describe the information in images with a lot of uncertainties.
- A vector in the eigenspace is directly mapped onto the controller output based on the B-spline model. This makes real-time computation possible.
- Designing the controllers is simple and identical for both low and high dimensional controllers. Both force and vision controllers are of the same type. The B-spline fuzzy controller can be trained in a straightforward manner because modification of control vertices only results in local change of the control surface.

In this approach no complex programming and knowledge about vision and force control is needed. We have shown that this approach is very promising for realizing efficient robot assembly skills based on sensorimotor coordinations.

References

[1] P. K. Allen, A. T. Miller, P. Y. Oh, and B. S. Leibowitz. Integration of visison and force sensors for grasping. In *Proc. IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems*, pages 349–, 1996.

[2] J. Baeten, Herman Bruyninckx, and J. De Schutter. Combining eye-in-hand visual servoing and force control in robotic tasks using task frame. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 1999.

[3] J. L. Clark and A. L. Yuille. *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers, 1990.

[4] C. Colombo, B. Allotta, and P. Dario. Affine visual servoing: A framework for relative positioning with a robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 464–471, 1995.

[5] M. Jägersand, O. Fuentes, and R. Nelson. Experimental evaluation of uncalibrated visual servoing for precision manipulation. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 2874–, 1997.

[6] A. Knoll, B. Hildebrandt, and J. Zhang. Instructing cooperating assembly robots through situated dialogues in natural language. In *IEEE International Conference on Robotics and Automation*, 1997.

[7] G. T. McKee. What can be fused? In J. K. Aggarwal, editor, *Multisensor Fusion for Computer Vision*, NATO ASI Series, pages 71–. Springer-Verlag, 1993.

[8] W. T. Miller. Real-time application of neural networks for sensor-based control of robots with vision. *IEEE Transactions on System, Man and Cybernetics*, 19:825–831, 1989.

[9] S. K. Nayar, H. Murase, and S. A. Nene. Learning, positioning, and tracking visual appearance. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3237–3244, 1994.

[10] B. J. Nelson and P. K. Khosla. Force and vision resolvability for assimilating disparate sensory feedback. *Transaction on Robotics and Automation*, 12(5):714–, 1996.

[11] Hugh F. Dur rant Whyte. *Integration, Coordination and Control of Multi-Sensor Robot Systems*. Kluwer Academic Publishers, 1988.

[12] Y. v. Collani, C. Scheering, Z. Zhang, and A. Knoll. A neuro-fuzzy solution for integrated visual and force control. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 1999.

[13] Y. v. Collani, Z. Zhang, and A. Knoll. A neuro-fuzzy solution for fine-motion control based on vision and force sensors. In *IEEE International Conference on Robotics and Automation*, 1998.

[14] B. H. Yoshimi and P. K. Allen. Active, uncalibrated visual servoing. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 156–161, 1994.

[15] J. Zhang and A. Knoll. Constructing fuzzy controllers with B-spline models – principles and applications. *International Journal of Intelligent Systems*, 13(2/3):257–286, February/March 1998.

[16] J. Zhang, Y. v. Collani, and A. Knoll. B-spline fuzzy controller to acquire sensor-based assembly skills. In *IEEE International Conference on Robotics and Automation*, 1997.