

# A Learning-Based Multisensor Fusion Approach for Fine Motion Control of Robot Arms

Dr. Jianwei Zhang, Dipl.-Inform. Yorck von Collani, Prof. Dr.-Ing. Alois Knoll  
Technical Computer Science, Faculty of Technology,  
University of Bielefeld, 33501 Bielefeld, Germany  
E-mail: zhang|ycollani|knoll@techfak.uni-bielefeld.de

## 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 multisensor 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.

## 1 Introduction

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 have awoken great interest. Conventional techniques try to exploit a common representation space to achieve a fused model of the environment [3]. 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 [6] force and vision feedback is combined using so called *vision and force resolvabilities*. In another approach presented in [2], 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. The commonness to nearly all these approaches is an explicit modelling of the sensor properties in order to combine the information.

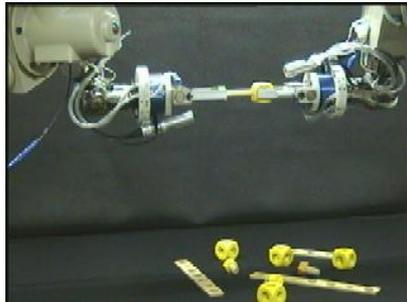
CMAC neural networks may tackle the problem of dimensionality. In [4] twelve inputs represent four joint positions of the robot, four image parameters and their desired changes. The outputs are the control signals for the four robot joints. In [5] learning of vision-based positioning based on visual 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 one-dimensional 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. In this work we fuse visual information from two uncalibrated cameras as well as from one camera and a force/torque sensor. We do not use

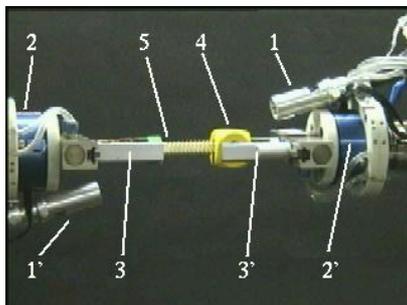
any explicit models but employ an adaptive neuro-fuzzy scheme to *learn* the appropriate robot motions necessary to perform a complex screw-task. We propose the following learning-based sensor fusion approach and apply it to a real robot system with two arms and multiple vision and force/torque sensors. These external sensors are used in parallel 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.

## 2 Problem Description

### 2.1 Experiment Setup



(a)



(b)

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

The problem scenario (Fig. 1), screwing

a screw (5) into a nut (4) with two cooperating robots (Puma 260), 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. Two global view cameras are installed: one as overhead, the other as side-view in front of the two arms.

### 2.2 Uncertainties

For a general-purpose arm/gripper system, the following two types of uncertainties must be taken into account: grasping precision and slippage of the part in the hand. These uncertainties can cause the following two concrete problems: a). the screw is not centrally grasped; and b). the screw is obliquely grasped (Fig. 2).

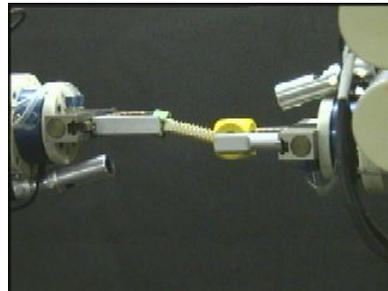


Figure 2: An inconvenient start-situation for screwing.

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 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.

### 3 A Neuro-Fuzzy Model for Vision-Based Control

#### 3.1 B-Spline Model

The controller for force control can be efficiently realised using the B-spline fuzzy controllers proposed in our earlier work [8, 9].

- 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 [8].

#### 3.2 Dimension Reduction

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.

#### 3.3 Projection into Eigenspace

A well-known technique for dealing with multivariate problems in statistics is the *principal component analysis* (PCA). As shown in [5], 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 [7].

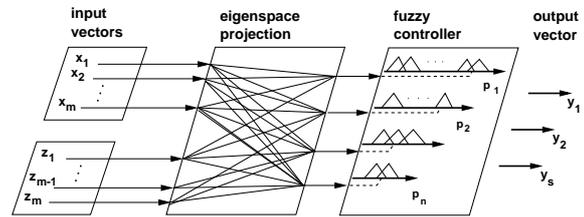


Figure 3: The structure of a fuzzy controller based on eigenspace projection by fusing images ( $x, z$ ).

An eigenvector, denoted as  $\vec{a}_i$ , is computed as  $[a_{1,i}, a_{2,i}, \dots, a_{m,i}]^T$ . The eigenvectors form an orthogonal basis for representing the original individual sensor patterns. Assume that the eigenvectors  $\vec{a}_1, \vec{a}_2, \dots$  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.  $\vec{a}_i$  defines the  $i$ th dimension in the eigenspace. The projection of an input vector  $\vec{x} = [x_1, x_2, \dots, x_m]^T$  onto eigenvector  $\vec{a}_i$ , called the  $i$ th principal component, is  $p_i = a_{1,i}x_1 + a_{2,i}x_2 + \dots + a_{m,i}x_m$ . The complete projection can be represented as:  $[\vec{a}_1, \dots, \vec{a}_n]^T \cdot \vec{x} = [p_1, \dots, p_n]^T$ .

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. 3.

## 4 Implementation

In this work we implemented two different controllers to adjust the orientation of the screw after contact is established:

1. Images from two different cameras are fused to determine the orientation of the screw.
2. Information of a force/torque sensor and the related camera are fused.

In both tasks we use the same B-spline neuro-fuzzy model [8].

For dimension reduction, we employ a method that extracts automatically the needed features from one or two fused images to compensate the uncertainties. If the image of a camera is regarded as a vector, this high-dimensional sensor image is obviously 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.

The following procedures are needed to implement a controller:

1. Sampling training data.
2. Calculating eigenvectors.
3. Training the fuzzy controller.

These procedures will be further described in our final paper.

## 5 Experimental Results

The vision-based controller is learned with 363 training images, shifting the screw

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

### 5.1 Fusing two cameras

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.

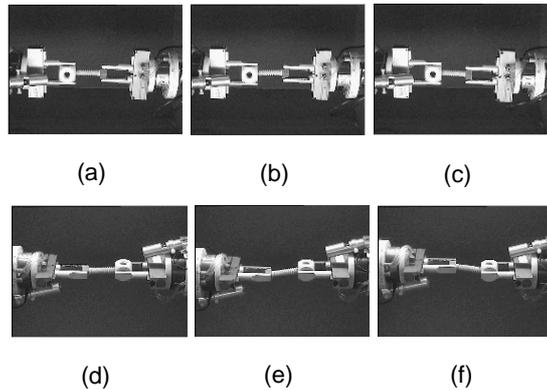


Figure 4: Typical images taken by the external cameras (a)–(c) viewpoint from above, (d)–(f) side view.

To correctly project the images into the pre-trained eigenspace, the “focus of interest” needs to be first selected. For the task to correct the screw orientation, rectangular sub-region of the images are clipped, Fig. 5.

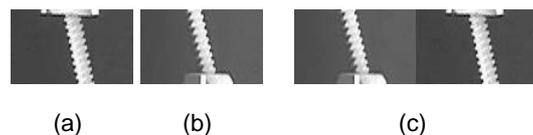


Figure 5: Clipped images (a) from camera 1 and (b) camera 2 and (c) the resulting merged image.

After merging the two images and pro-

	Mean square error [° <sup>2</sup> ]	Maximal error [°]	Worst case error [°]
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, maximal error and worst case error for angle around N- and O-direction.

jecting them into the eigenspace, we used the three largest eigenvectors as input for the B-spline fuzzy controller. Each eigenvector is covered with 10 B-splines as membership functions.

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.

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 maximal error of the test images around the N- and O-direction are smaller. The error never grows, Tab. 1.

## 5.2 Combining visual and force/torque information

Similar to the first task we also use the three greatest eigenvectors as input for the B-spline fuzzy controller. Additional the

torque around the N- and the O- axis of the tool is used as input. Each eigenvector is covered with 10 B-splines and the torque with 5 B-splines as membership functions. Detailed report on the mean square error of the controller, the maximal error and the largest worst case errors of the test images around the N- and O-direction with and without additional force will be given by the following table. This experiment shows that the fusion of vision and force/torque data produce better results in comparison with the unfused case.

	Mean square error [° <sup>2</sup> ]	Maximal error [°]	Worst case error [°]
N- direction			
one camera	8.62	7.35	7.07
camera + torque	6.6	8.75	–
O- direction			
one camera	12.9	11.64	7.65
camera + torque	8.18	8.2	–

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

## 6 Conclusions

We have shown that the B-spline model in combination with dimension reduction may be utilised for sensor fusion and high-dimensional problems such as visually guided fine-motion. We have implemented the approach with a two-arm robot system based on supervised off-line learning with input of the vision system and force/torque sensor.

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 realising 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 vision 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] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.