

Sensor Fusion Approaches for Observation of User Actions in *Programming by Demonstration*



M. Ehrenmann, R. Zöllner, S. Knoop and R. Dillmann
Universität Karlsruhe (TH)
Institute for Process Control & Robotics
Karlsruhe, D 76128, Germany
Email: {ehrenman|zoellner|knoop|dillmann}@ira.uka.de

Abstract

Good observation of a manipulation presentation performed by a human teacher is crucial to further processing steps in Programming by demonstration (Pbd) which has reached a prime of importance in interactive robot programming. This paper outlines a sensor fusion concept for hand action tracking observing hand posture, position and applied forces. As input sources serve on the one hand a data glove which will classify several gestures and grasps, a stereo camera head and several force sensors mounted on the finger tips. The hardware used at our institute is presented as well as first implementations of measurement and fusion approaches. Accuracies of first experiments are also given.

1 Introduction

Pbd addresses the high cost of robot program development. Usually this is done by skillful experts because of the use of advanced sensor systems and the existence of strong requirements with respect to the robot's flexibility. These skills may exist in industrial environment but they will certainly not be feasible when personal robots are integrated in hospitals, environments for handicapped persons or at home. For opening the new, mainly consumer-oriented service robot market, it is therefore essential to develop techniques that allow untrained users to use such a personal service robot both safely and efficiently.

The aim of *Pbd* is to let arbitrary persons program robots by simply giving a demonstration of how to solve a certain task to a sensor system and then have a system interpret his actions and map them to a specific

manipulator. Although detecting and understanding the user's actions and intentions turned out to be a quite difficult task, the benefits of intuitive robot programming should not be underestimated.

In this paper an approach for a sensor system with cameras and a data-glove will be presented. Section 2 will give a brief overview on today's *Pbd* techniques. In section 3 the *Pbd*-system currently running at our institute and the employed sensor devices will be outlined as well as the implemented approaches. Section 4 focuses on sensor fusion between the vision system and magnetic tracker which is used in order to track movements followed by the presentation of force processing in section 5 which serves for better grasp analysis.

2 State of the art

Realization of recognition and interpretation of continuous human action sequences is critical to *Pbd*. Though, there are few publications regarding sensors including visual processing. Kuniyoshi et al. [14, 15] presented a system with a visual hand-tracker module that is able to detect grips and drops of objects. However, only one type of grasping is classified and the hand is constrained to appear under a certain angle. Kang [11] used a data glove in combination with depth images computed from recorded image sequences for a reconstruction of what has been done. Depth images are yield by the projection of structured light thus undergoing real-time constraints.

Since elementary operations consist of movements, a lot effort has been spent tracking and reconstructing the trajectories of objects [24, 25], a robot's effector [17] or user's hand [18, 8, 27, 19]. Some authors consider demonstrations only in a virtual or augmented

environment [21].

Many researchers are interested in the recently raising gesture and grasp recognition field. Today's grasp detectors regard contact points between hand and objects in order to classify a grasp [12] or the hand posture itself [10]. Mostly, static grasps are considered. Gestures are used in order to command a robot and tracked visually [23, 1, 13]. These results can be of use for *Pbd* also: hand configurations can be used to direct a gripper [16] or trigger implemented skills [22]. This way, *Gesture based programming* has gathered some importance [26].

3 Experimental setup and prior work

Focusing service tasks in households and workshop environments, for *Pbd* information about grasping states, movements, forces and objects is needed. Therefore, we consider combining results of as many suiting sensor types as possible in order to obtain as much information as possible from a single demonstration.

3.1 Sensors

As sensors for observing a user demonstration of a manipulating task, a VPL data glove, a camera head, a Polhemus magnetic tracker and force sensors both mounted on the glove are used in a fixed rack (see figure 1).

Because of its many degrees of freedom and changing of shape, it is very difficult to extract posture information about a user's hand solemnly out of image sequences. Especially information about its particular grasping state is hard to obtain. Following [20], we consider data gloves as good sensors for obtaining this kind of information. In order to record a demonstration trajectory, all the VPL data glove sensor data is used while the measurements of the Polhemus tracker are fused with visual tracking data. Unfortunately, the magnetic tracker is accurate only in a small region around the field emitter due to metallic objects in our laboratory.

Visual tracking follows a marker fixed on the magnetic tracker. The camera head employs three greyscale Pulnix TM765i cameras and AMTEC turn and tilt modules. For grabbing, a Matrox Genesis frame grabber is used on a standard PC. Additionally, visual data is used for determining the types of manipulable objects and positions.

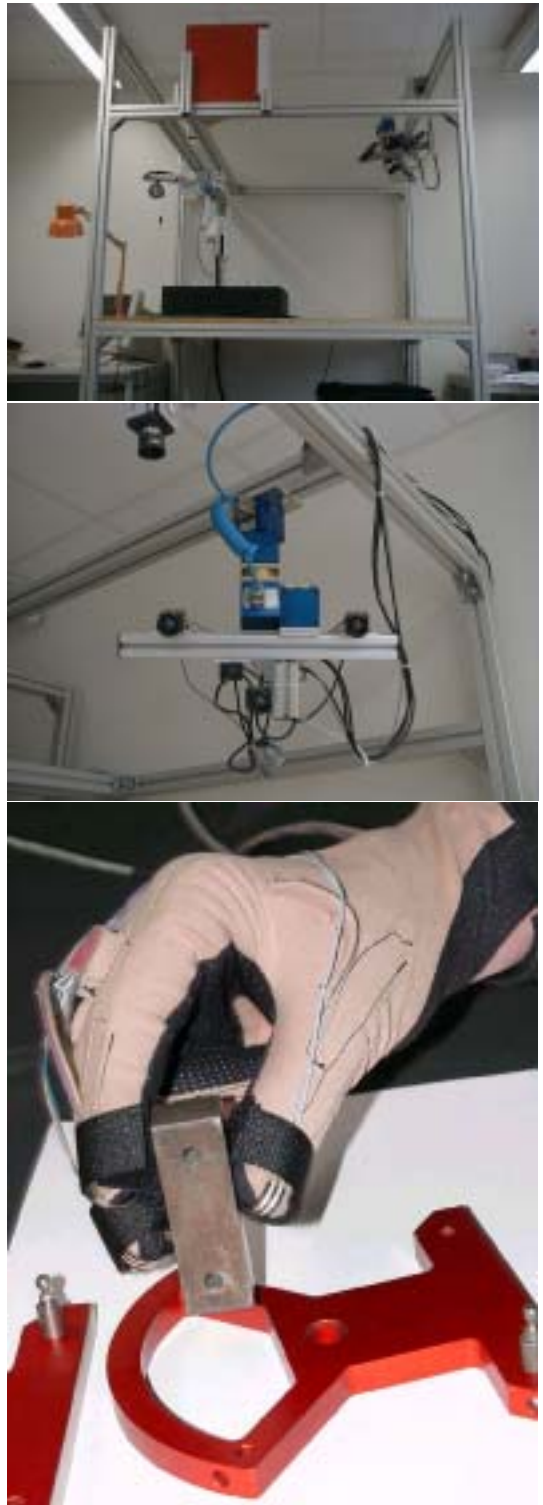


Figure 1: Experimental environment and used sensor devices. From top to bottom: demonstration rack, camera head and data glove with fixed force sensors.

3.2 Pbd Approach

According to the *Pbd* cycle presented in [5], we first check for objects present in the scene that a user is about to manipulate. This is done via the camera head using state of the art image processing methods [9, 4]. After reconstructing their particular positions in the rack, the user's hand is being tracked recording the trajectory given by the magnetic and visual tracker. The recorded trajectory is then analyzed, interpreted and mapped to a manipulator (see [3, 6]).

Regarding the analysis of the demonstration, we have shown that a grasp can be detected and classified according to the Cutkosky hierarchy [2] with high precision and robustness by a neural network classifier [7]. We use this information combined with movement speed considerations to determine grasp events and movements. Since trajectories are stored with respect to the manipulated objects, it is easy to map the movements to robots in a similar environment.

4 Sensor Fusion for Tracking: Visual and Magnetic Trackers

Movements have to be tracked as the basis of a demonstration. Vision and magnetic trackers do have their particular drawbacks doing this. In order to get the most out of it, their measurements have to be fused.

4.1 Accuracies and Measurement Rates

Calibration for the vision system is done with 200 reference points via DLT matrices resulting in an accuracy of $\approx 3\text{mm}$ s all over the demonstration area in the rack. The data glove on the other hand gives very accurate measurements ($\approx 2\text{mm}$ s). Unfortunately, this applies only to a small region around the emitter ($\approx 60\text{mm}^3$, see figure 2). Iron parts in our laboratory disturb the magnetic field making the results completely useless outside these bounds.

Furthermore, due to magnetic fields inherent in the lab, the Polhemus sensor gives translated values and has to be calibrated before usage.

Visual tracking uses a marker object (see figure 3) that is segmented via binarizing the image with an adaptive threshold and then selecting three blobs that meet several conditions, e.g. size, compactness and existence of a hole. Thus, tracking is very stable even with unstructured background and very fast ($> 35\text{Hz}$). This enables the head to follow even rapid movements.

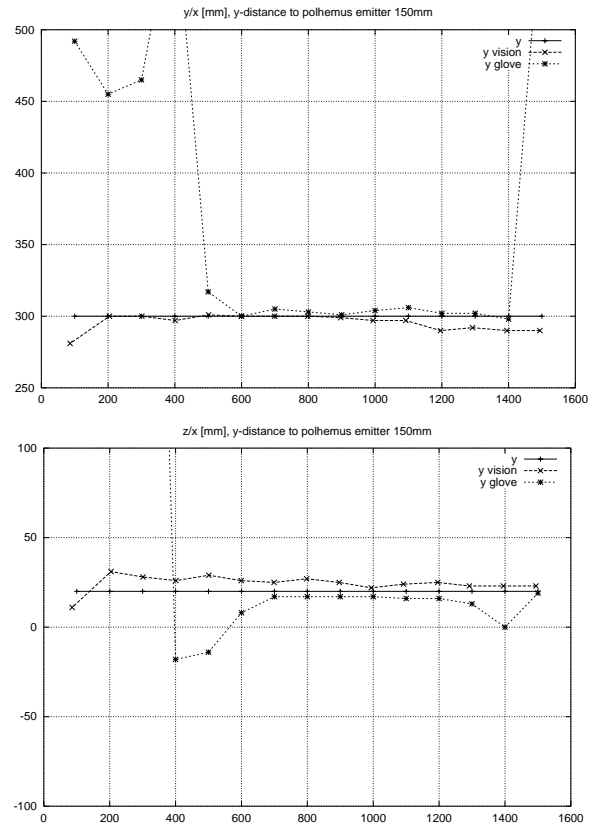


Figure 2: X/Y and X/Z plane with vision and data glove measurements. The field emitter is located at $x = 1040$. Polhemus values farther than 30mm s are highly disturbed.



Figure 3: Marker object for tracking.

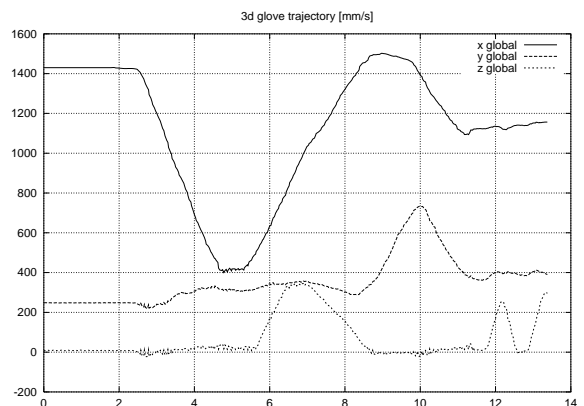


Figure 4: X , Y and Z value of a sample trajectory recorded by the vision sensor.

As can be seen in figure 4, there is few noise (except for high accelerations at second 3) and the tracker rarely loses the target. The maximum marker velocity depends on the distance between cameras and user hand; at 1.5m we could move the target at more than $0.5 \frac{m}{s}$. On the other hand, the data glove is read out at 25Hz and cannot lose the target.

4.2 Fusion Approaches

At the moment, the Polhemus tracker values are used for initializing the camera head. The cameras serve for calibrating the magnetic sensor. Afterwards only the values obtained by visual tracking are used. Only in case the marker is lost the glove position is requested and the head positions itself to the according coordinate calculating the inverse kinematics.

This approach is not sufficient because of two reasons:

- Rotational coordinates given by the Polhemus sensor are more accurate because of the small marker size. Furthermore, the magnet tracker gives good values close to the emitter.
- Steps occur when switching from one sensor source to the other.

Thus, not only sensors but sensor data have to be combined. Two problems arise facing this question:

- Weight factors have to be found determining the certainty of the particular values:

Data glove: Error distribution can not be modelled as white noise because values depend on the magnetic field of the rack. At a certain distance from the emitter, values are completely incorrect. That is why heuristic methods have to be used (thresholds, confidence intervals).

Image processing: Error distribution can simply be modelled linearly. Variance of the measurement error depends on the distance between camera and target as well as on marker size, CCD resolution and acceleration.

- Measurements are not equidistant in time nor received at the same time stamp. In order to combine the received measurements, samples have to be interpolated to the same frequency and time stamps. Thus, sensors have to be synchronized.

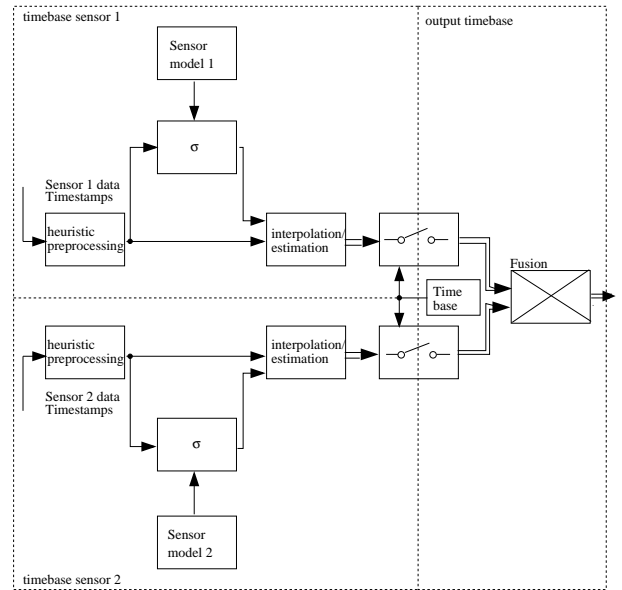


Figure 5: Fusion algorithm to be implemented.

We propose a sensor fusion approach following the depicted structure (see figure 5). Before combining the raw data, time bases of both sensor types are mapped to the output time base. In a first stage, the measurements are tested for plausibility and steps are filtered out. A sensor model based function calculates the corresponding error covariance. For both sensors, these values and the measurements get sampled with the same frequency. When using an estimator instead of simple interpolation, noise filters may be adapted for the particular sensor types. Fusion itself is a weighted sum of the results. A last filter smoothes steps that may occur due to weight changes or erroneous measurements (not included in the figure).

5 Grasping forces

This section gives a brief overview of the integration of tactile sensors in the data glove in order to perform a better grasp recognition. One of the lacks of the above described PbD system is the accurate determination of grasp and ungrasp actions. To improve this tactile sensors were attached on the fingertips of the data glove, as shown in figure 1. The active surface of the sensors is covering the hole fingertips. The wires to the interface device are conducted on the upper side of the fingers, allowing the user to move his finger with maximal agility.

5.1 Sensors Properties

For a first approach low-price, industrial sensors shipped by Interlink company were used. They are based on an Force Sensing Resistor (FSR). For our application a circular layout with one cm diameter of the active surface was selected.

When applying an increasing force to the sensors active surface the resistance decreases. The FSR response approximately follows an inverse power-law characteristic ($U \approx 1/R$). For a force range of 1–100N the sensor characteristics are good enough for detecting grasp actions. This range shows a hysteresis below 20% and the repeatability of measurements is around $\pm 10\%$. Following these restrictions the force is quantised into 30 – 50N units.

Some remarks for the use of the sensor have to be made. The active surface is very sensitive concerning bending ($r < 2.5mm$), since it can cause tenseness in the material. This may result in pre-loading and false readings. Therefore we applied the active surface on an thin and rigid plate. This way, good and reliable results are achieved. Whether this configuration shows little drift of readings when static forces are applied.

5.2 Integrating Force Results in the PbD Cycle

Our PbD system presented in [5], consists of following phases: *Observation*, *Trajectory segmentation*, *Mapping the trajectory segments to elemental Operations* and *Model based mapping*. The first phases are dealing with the registration of the user’s demonstration and intention. For manipulation tasks the recognition of contact between hand and object, is to be performed in order to segment the trajectory. Evidently this is easily obtained from the force values with a threshold based algorithm.

To improve the reliability of the system the results of this algorithm are merged with the values obtained by older, previous implemented recognition routines. These are based on the analysis of trajectories of finger poses, velocity and acceleration w.r.t. to minima. Figure 6 shows the trajectories of force values, finger joint an velocity values of three *Pick & Place* actions.

For the *Model based mapping* gathered force values are used to determine the right grasp type to map to. So, the grasp type is defined by the contact points and the applied force. The fact that the sensors show a drift on statical strain implies that the gained information is rather qualitative than quantitative. Due to this fact ten classes can be defined for characterizing the grasping force.

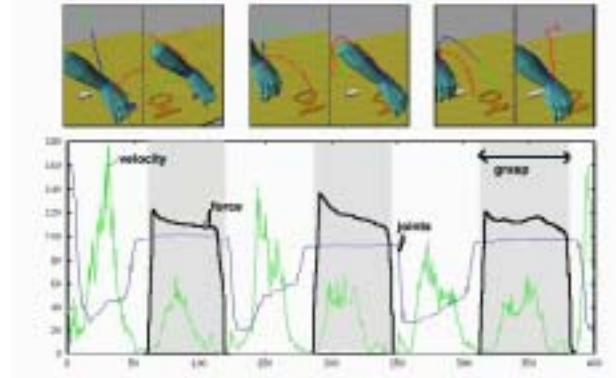


Figure 6: Analyzing segments of a demonstration: force values and finger joint velocity

Future works will analyze force characteristics with respect to grasp type, weight, surface features of the grasped object and trajectory type in order to extract further significant information. This will be integrated in the systems knowledge base and used for task recognition or mapping the demonstrated actions to a robot system.

Acknowledgement

This work has partially been supported by the BMBF project “Morpha”. It has been performed at the Institute for Real-Time Computer Systems & Robotics, Department of Computer Science, University of Karlsruhe.

References

- [1] A. Blake and M. Isard. *Active Contours*. Springer, 1998.
- [2] M. R. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation*, 5(3):269–279, 1989.
- [3] R. Dillmann, O. Rogalla, M. Ehrenmann, R. Zöllner, and M. Bordegoni. Learning robot behaviour and skills based on human demonstration and advice: the machine learning paradigm. In *9th International Symposium of Robotics Research (ISRR 99)*, pages 229–238, Snowbird, Utah, USA, 9.-12. Oktober 1999.
- [4] M. Ehrenmann, D. Ambela, P. Steinhaus, and R. Dillmann. A comparison of four fast vision based object recognition methods for programing by demonstration applications. In *Proceedings of the 2000 In-*

- ternational Conference on Robotics and Automation (ICRA), volume 1, pages 1862–1867, San Francisco, Kalifornien, USA, 24.–28. April 2000.
- [5] M. Ehrenmann, P. Steinhaus, and R. Dillmann. A multisensor system for observation of user actions in programming by demonstration. In *Proceedings of the IEEE International Conference on Multi Sensor Fusion and Integration (MFI)*, volume 1, pages 153–158, Taipeh, Taiwan, August 1999.
- [6] H. Friedrich. *Interaktive Programmierung von Manipulationssequenzen*. PhD thesis, Universität Karlsruhe, 1998.
- [7] H. Friedrich, V. Grossmann, M. Ehrenmann, O. Rogalla, R. Zöllner, and R. Dillmann. Towards cognitive elementary operators: grasp classification using neural network classifiers. In *Proceedings of the IASTED International Conference on Intelligent Systems and Control (ISC)*, volume 1, Santa Barbara, Kalifornien, USA, 28.–30. Oktober 1999.
- [8] D. Gavrila and L. Davis. Towards 3d model-based tracking and recognition of human movement: a multi-view approach. In *International Workshop on Face and Gesture Recognition, Zürich, 1995*.
- [9] J. González-Linares, N. Guil, P. Pérez, M. Ehrenmann, and R. Dillmann. An efficient image processing algorithm for high-level skill acquisition. In *Proc. of the International Symposium on Assembly and Task Planning (ISATP), Porto, Portugal*, pages 262–267, Juli 1999.
- [10] H. Hashimoto and M. Buss. Skill acquisition for the intelligent assisting system using virtual reality simulator. In *Proceedings of the 2nd International Conference on Artificial Reality and Tele-existence, Tokyo, 1992*.
- [11] S. Kang. *Robot Instruction by Human Demonstration*. PhD thesis, Carnegie Mellon University, Pittsburg, Pennsylvania, 1994.
- [12] S. Kang and K. Ikeuchi. Toward automatic robot instruction from perception: Mapping human grasps to manipulator grasps. *Robotics and Automation*, 13(1):81–95, Februar 1997.
- [13] H. Kestler, M. Borst, and H. Neumann. Einfache Handgestikerkennung mit einem zweistufigen Nearest-Neighbour Klassifikator. Technical report, Universität Ulm, SFB 527, 96/6, 1996.
- [14] Y. Kuniyoshi, M. Inaba, and H. Inoue. Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *IEEE Transactions on Robotics and Automation*, 10, 1994.
- [15] Y. Kuniyoshi and H. Inoue. Qualitative recognition of ongoing human action sequences. In *13th International Joint Conference on Artificial Intelligence, 1993*.
- [16] D. Mostafa. Anthropomorphic interface for robot arm programming through a data glove. In *IEEE International Symposium on Industrial Electronics (ISIE)*, pages 326–328, 1994.
- [17] M. Päsche and J. Pauli. Vision based learning of gripper trajectories for a robot arm. In *International Symposium on Automotive Technology and Automation (ISATA), Florence*, pages 235–242, 1997.
- [18] J. Rehg and T. Kanade. Visual tracking of high DOF articulated structures: an application to human hand tracking. In *ECCV*, pages 35–46, 1994.
- [19] N. Shimada and Y. Shirai. 3d hand pose estimation and shape model refinement from a monocular image sequence. In *Proceedings of the VSMM, Gifu*, pages 423–428, 1996.
- [20] D. Sturman and D. Zeltzer. A survey on glove-based input. *IEEE Computer Graphics and Applications*, 14(1):30–39, 1994.
- [21] K. Tanaka, N. Abe, M. Ocho, and H. Taki. Registration of virtual environment recovered from real one and task teaching. In *Proceedings of the IROS 2000, Seoul, Korea, 2000*.
- [22] J. Tatsuno, S. Matsuyama, Y. Kokubo, K. Kawabata, and H. Kobayashi. Human friendly teaching for industrial robots. In *IEEE International Workshop on Robot and Human Communication, 1996*.
- [23] J. Triesch and Chr. von der Malsburg. Robotic gesture recognition. In *Proceedings of the Bielefeld Gesture Workshop*. Springer, 17.-19. September 1997.
- [24] A. Ude. *Rekonstruktion von Trajektorien aus Stereobildfolgen für die Programmierung von Roboterbahnen*. PhD thesis, Universität Karlsruhe, 1996. Erschienen in: VDI Verlag, Fortschr. Ber. VDI Reihe 10 Nr. 448. Düsseldorf.
- [25] A. Ude. Filtering in a unit quaternion space for model-based object tracking. *Robotics and Autonomous Systems*, 28:163–172, 1999.
- [26] R. Voyles, J. Morroy, and P. Khosla. Gesture-based programming for robotics: Human augmented software adaption. *IEEE Intelligent Systems*, pages 22–29, November/Dezember 1999.
- [27] M. Yamamoto and K. Koshikawa. Human motion analysis based on a robot arm model. In *CVPR*, pages 664–665, 1991.