Dynamic Grasp Recognition within the Framework of Programming by Demonstration

R. Zöllner, O. Rogalla, R. Dillmann
Universität Karlsruhe (TH)
Institute for Process Control & Robotics
Karlsruhe, D 76128, Germany
Email: {zoellner|rogalla|dillmann}@ira.uka.de

J.M. Zöllner
Forschungszentrum Informatik (FZI)
Karlsruhe, D 76133, Germany
Email: zoellner@fzi.de

Abstract

Programming robots by unexperienced human users require methods following the Programming by Demonstration (PbD) *paradigm. The main goal of these systems is to allow the unexperienced human user to easily integrate motion and perception skills or complex problem solving strategies. Unfortunately actual* PbD *systems are dealing only with manipulations based on* Pick & Place *operations. This paper describes how fine manipulations like detecting screw moves can be recognized by a* PbD *system. Therefore the question:* "What happens during a grasp ?" *has to be answered. In order to do this, finger movements and forces on the fingertips are gathered and analyzed while a object is grasped. This assume vast sensory employment like a data glove and integrated tactile sensors. An overview of the used tactile sensors and the gathered signals is given. Furthermore a classification of the recognized* Dynamic Grasp *is pointed out as well as the classification method based on a* Support Vector Machine (SVM).

1 Introduction

Using personal and service robots implies high demands on the programming interface. The interaction of these robots with humans and there programming needs the developing of new techniques that allow untrained users to use such a personal service robot both safely and efficiently. *PbD* is one way to meet these high requirements.

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. Learning systems are needed capable of extracting knowledge from watching users demonstration. Heterogeneous sensor inputs like vision, tactile or position information are required of such systems.

This paper presents an approach of a *PbD* system, which handles more than only *Pick & Place* operations. In order to detect fine manipulations, a grasp is analyzed with respect to finger movements and forces performed on the fingertips. 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 integration of tactile sensors in a data glove in order to detect contact phases during a users demonstration. The reliable detection of grasped and ungrasped fragments is crucial for analyzing a given demonstration. Section 5 performs a analyze of the force sensor signals in order to divide the grasp in segments. Finally section 6 presents and classify dynamic grasps. The classification of the dynamic grasp is done through a *Support Vector Machine*, by using a time delay approach.

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. [18, 19] 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 [15] 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 [28, 29], a robot's effector [22] or user's hand [23, 9, 32, 24]. Some authors consider demonstrations only in a virtual or augmented environment [27].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 [16] or the hand posture itself [11]. Mostly, static grasps are considered.

In the domain of recording tactile information many tactile sensors have been developed in the past ten and more years. Some good surveys of tactile sensing technologies were provided by Nicholls et all [21] and Howe [13]. Several researchers have used tactile feedback for determining object shapes or force primitives from users demonstration [1, 2, 17, 31]. Most of these works are trying to map the extracted force characteristics directly to robot actions [25, 12, 20].

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 [26], 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 merged with visual tracking data.

Figure 1: Experimental environment - demonstration rack and data glove with mounted tactile sensors.

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.

3.2 *PbD* Approach

According to the *PbD* cycle presented in [6], 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 [10, 5]. 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 [4, 7]). So far only *Pick&Place* operations were considered.

Regarding the analysis of the demonstration, we have shown that a static grasp can be detected and classified according to the Cutkosky hierarchy [3] with high precision and robustness by a neural network classifier [8]. We used this information combined with movement speed considerations to determine grasp events and movements. The next section shows how this segmentation step is extended by using tactile sensors.

4 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.

4.1 Sensors Properties

For a first approach low-price, industrial sensors of the Interlink company, based on an Force Sensing Resistor (FSR) were used. For our application an circular layout with one cm diameter of the active surface was selected (see figure 2).

Applying a 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 1-100 N the sensor characteristics are good enough for detecting grasp actions. This range shows a hysteresis below 20% and the repeatability of measurements is around ±10%. Following these restrictions the force is quantized into $30 - 50N$ units.

Figure 2: Tactile Sensor

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. Proceeding so, good and reliable results are achieved. Whether this configuration shows little drift of readings when static forces are applied.

4.2 Integrating Force Results in the *PbD* Cycle

The main benefit from gathering force values is assigned to *Trajectory segmentation* of the users demonstration. 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 analyze of trajectories of finger poses, velocity and acceleration w.r.t. to minima. Figure 3 shows the trajectories of force values, finger joint an velocity values of three *Pick&Place* actions.

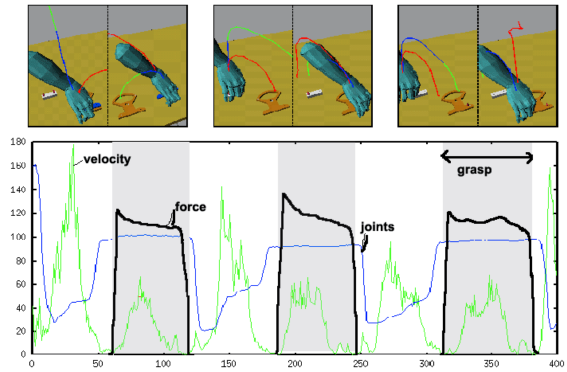

Figure 3: Analyzing segments of a demonstration: force values and finger joint velocity.

5 Analyze of a Grasp

While the last section described how force sensors can be used to segment users demonstration in *Pick & Place* fragments, the aim of this section is to analyze what happens while a object is grasped. Figure 3 shows that the shape of the force graph features a relative constant plateau. Due to the fact that no external forces are applied to the object this effect is plausible. But if the grasped object collides with the environment the force profile will chance. The results are high peaks i.e. both amplitude and frequency are rising very fast (see figure 4).

Empirical test have shown that at least three different profiles can be distinguished:

- Static Grasp
  Here the gathered force values are nearby constant. The force profile shows characteristic plateaus, where the height points out the weight of the grasped object.

- External Forces

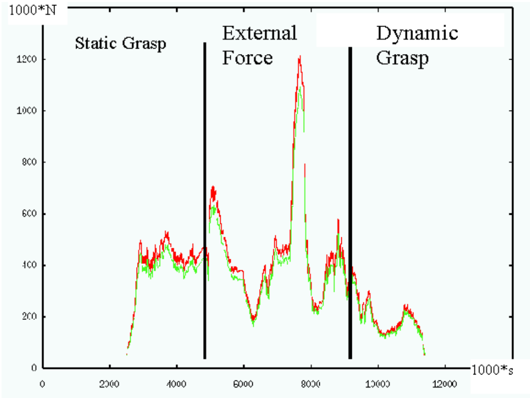

Figure 4: Variation of force signals during a Grasp

The force graph of this class shows high peaks. Because of the hysteresis of the sensors no quantitative prediction about the applied forces can be made. A proper analyze of external forces applied to a grasped object will be subject of further works.

- Dynamic Grasps
  During a dynamic grasp the both amplitude and frequency oscillate moderate, as a result of finger movements performed by the user.

Next section will point out what dynamic grasps are and how they can be classified.

6 Classification of Dynamic Grasps

For describing various household activities like opening a twisted cap or screwing a bold in a nut, simple operation like *Pick & Place* are not longer adequate. Therefor new operations like *Dynamic Grasps* need to be included in the *PbD* system.

6.1 Dynamic Grasps

With *Dynamic Grasps* we denote operations like screw, insert etc. which all have in common that finger joints are changed during a object is grasp (i.e. the force sensors provide non zero values). In our first approach we distinguish three elementary *Dynamic Grasps*:

- Screw
  This operation describes rotations around the z-axes (see figure 5) which is performed when screwing a bold.
- Twist
  When opening a twisted cap a *Twist Grasp* can be performed. It denotes the rotation around the x-axes referred in figure 5.
- Insert
  Other than the upper two dynamic grasps, the *Insert Grasp* specifies a translatoric move along the z-axes like shown in figure 5.

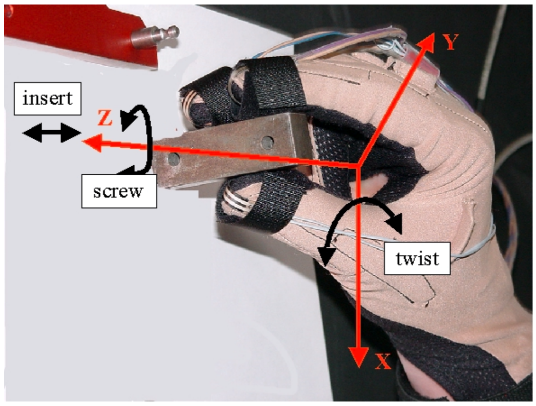

Figure 5: Directions of the elementary dynamic grasps

These three elementary dynamic grasps lead to various characteristics like number of finger which where involved in the grasp (i.e. form 2 to 5) and the direction in which the rotation or translation was performed. On the another hand elementary dynamic grasps can be combined to an complex grasp. For example during a *screw operation* a translatoric component i.e. a *insert operation* can be performed. Next section gives a brief overview about SVM's, before the results are presented in the last section.

6.2 Support Vector Machine Classifier

Support vector machines are a general class of statistical learning architectures, which are rising up with a profound theoretical foundation as well as excellent empirical performance in a variety of applications. Originally developed for pattern recognition, the SVMs justify their application by a large number of positive qualities, like: fast learning, accurate classification and in the same time a high generalization performance. The basic training principle behind the SVM is to find the optimal class-separating hyperplane so, that the expected classification error for unseen examples is minimized. Using the *kernel trick* and the implicit transformation in a high dimensional working space, leads to nonlinear separation of the feature space. The decision function becomes a linear combination of kernels of the training data:

$$f(x) = \sum_j \alpha_j y_j K(x, x_j) + b$$

where x_j are the training vectors with their corresponding labels y_j, and α_j are the Lagrange - multiplier. By performing the Lagrange - optimization for finding the optimal separating hyper plane just a small set of multiplier α are carried out as nonzero. The corresponding data points are the so-called support vectors. [30]

6.3 Experimental Results

For training the SVM Gaussian kernel functions, an algorithm based on the SVMLight [14] and the one-against-one strategy have been used. Three classes corresponding to the elementary dynamic grasps (i.e. screw, twist and insert executed in only one direction) where trained . Because of the fact that a dynamic grasp is defined by a progression of joint values a time delay approach was chosen. Consequently the input vector of the SVM Classifier comprised 50 joint configurations of 22 joint values. The training data set contained 294 input vectors. This data set is not enough to be representative, it shall only illustrate that this approach works. The results presented in Figures 6 and 7 and the fact that SVM's can learn from significant less data than neuronal networks, assure that this approach will work very well. For the final paper we will present results made with vast data sets.

Data	SVM1 ($\gamma = 0.001$)			SVM2 ($\gamma = 0.01$)		
Size	#SV	good	bad	#SV	good	bad
50x22	49	100%	0%	98	100%	0%
10x20	61	96,4%	3,5%	51	96,4%	3,5%

Figure 6: Classification with two SVM

Figure 6 shows results of two SVM's with different $\gamma's$ in order to test the generalization behavior. Where γ is indirect proportional to the squared variance of the Gaussian kernel function. Remarkable is the fact that the SVM1 needs only 49 support vectors (SV) for generalizing over 296 vectors i.e. 16,5% of the data set. Less number of SV improves not only the generalize behavior but also the runtime of the resulting algorithm during the application. The input vector of the second row contains only 10 joint configurations (of 20 joints). It was obtained by taking each fifth joint configuration, in order to shorten the input vector.

For a supplementary validation of the SVM's a new data set of 100 input vectors containing 53 correct labeled data (i.e. elementary dynamic grasps executed in the right direction) and 47 incorrect labeled data (elementary dynamic grasps executed in wrong direction) was used. As presented in figure 7 the SVM separates correct the *good* from the *bad* data according to the test data.

Data	SVM1 ($\gamma = 0.001$)		SVM2 ($\gamma = 0.01$)	
Size	good	bad	good	bad
50x22	53%	47%	53%	47%
10x20	45%	55%	47%	53%

Figure 7: Validation on a mixed data set with 53 correct labeled and 47 incorrect labeled dynamic grasps

7 Conclusion

This paper pointed out how a *PbD* system handling *Pick & Place* manipulation is enhanced by gathering Dynamic Grasps. In this context tactile sensors were mounted in a data glove in order to improve the reliably of detecting grasps in a user demonstration. Further on it was shown how Dynamic Grasps can be marked out by analyzing the force signals. Finally a new time delay approach based on a Support Vector Machine was realized in order to classify Dynamic Grasps.

Acknowledgment

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] P. Akalla, R. Siegwart, and M.R. Cutkosky. Manipulation with soft fingers: Contact force control. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, volume 2, pages 652–657, 1991.

[2] P. Berkelman and R. Hollis. Interacting with virtual environments using a magnetic levitation haptic interface. In *Proceedings of the 1995 IEEE/RSJ Intelligent Robots and Systems Conference*, Pittsburgh, PA, August 1995.

[3] 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.

[4] 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.

[5] 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 International Conference on Robotics and Automation (ICRA)*, volume 1, pages 1862–1867, San Francisco, Kalifornien, USA, 24.–28. April 2000.

[6] M. Ehrenmann, P. Steinhaus, and R. Dillmann. A multisensor system for observation of user actions in programing 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.

[7] H. Friedrich. *Interaktive Programmierung von Manipulationssequenzen*. PhD thesis, Universität Karlsruhe, 1998.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] S. Hirai and H. Asada. A model-based approach to the recognition of assambly process states using the theory of polyhedral convex cones. In *Proceedings of the 1990 Japan USA Symposium on Flexible Automation*, pages 809–816, Kyoto, Japan, 1990.

[13] R.D. Howe. Tactile sensing and control of robotic manipulation. In *Journal of Advanced Robotics*, pages pp 245–261, 1994.

[14] T. Joachims. Making large scale SVM learning practical. In B. Sch¨llkopf, C.J.C. Burges, and A.J. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169 – 184. Cambridge, MA, MIT press edition, 1999.

[15] S. Kang. *Robot Instruction by Human Demonstration*. PhD thesis, Carnegie Mellon University, Pittsburg, Pennsylvania, 1994.

[16] 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.

[17] D.A. Kontarinis, J.S. Son, W. Peine, and R.D. Howe. A tactile shape sensing and display system for teleoperated manipulation. In *Proceedings of the 1995 IEEE International Conference on Robotics and Automation*, volume 1, pages 641–646, May 1995.

[18] 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.

[19] Y. Kuniyoshi and H. Inoue. Qualitative recognition of ongoing human action sequences. In *13th International Joint Conference on Artificial Intelligence*, 1993.

[20] B.J. McCarragher. Force sensing from human demonstration using a hybrid dynamical model and qualitative reasoning. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, volume 1, pages 557–563, San Diego, May 1994.

[21] H.R. Nicholls and M.H. Lee. A survey of robot tactile sensing technology. In *International Journal of Robotics Research*, pages pp 3–30, June 1989.

[22] M. Päschke 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.

[23] J. Rehg and T. Kanade. Visual tracking of high DOF articulated structures: an application to human hand tracking. In *ECCV*, pages 35–46, 1994.

[24] 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.

[25] M. Skubic, S.P. Castriani, and R.A. Volz. Identifying contact formations from force signals: A comparison of fuzzy and neural network classifiers. In *IEEE 1997*, volume v. 8, pages 1623–1628, 1997.

[26] D. Sturman and D. Zeltzer. A survey on glove-based input. *IEEE Computer Graphics and Applications*, 14(1):30–39, 1994.

[27] K. Tanaka, N. Abe, M. Ooho, and H. Taki. Registration of virtual environment recovered from real one and task teaching. In *Proceedings of the IROS 2000, Seoul, Korea*, 2000.

[28] 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.

[29] A. Ude. Filtering in a unit quaternion space for model-based object tracking. *Robotics and Autonomous Systems*, 28:163–172, 1999.

[30] V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, Inc., 1998.

[31] R.M. Voyles, G. Fedder, and P.K. Khosla. Design of a modular tactil sensor and actuator based on an electrorheological gel. In *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, volume 4, pages 13–17, April 1996.

[32] M. Yamamoto and K. Koshikawa. Human motion analysis based on a robot arm model. In *CVPR*, pages 664–665, 1991.