Human-to-Robot Mapping of Grasps

Javier Romero Hedvig Kjellström Danica Kragić
Computational Vision and Active Perception Lab
Centre for Autonomous Systems
School of Computer Science and Communication
KTH, SE-100 44 Stockholm, Sweden
jrgn,hedvig,dani@kth.se

***Abstract*— We are developing a Programming by Demonstration (PbD) system for which recognition of objects and pick-and-place actions represent basic building blocks for task learning. An important capability in this system is automatic visual recognition of human grasps, and methods for mapping the human grasps to the functionally corresponding robot grasps. This paper describes the grasp recognition system, focusing on the human-to-robot mapping. The visual grasp classification and grasp orientation regression is described in our IROS 2008 paper [1]. In contrary to earlier approaches, no articulated 3D reconstruction of the hand over time is taking place. The input data consists of a single image of the human hand. The hand shape is classified as one of six grasps by finding similar hand shapes in a large database of grasp images. From the database, the hand orientation is also estimated. The recognized grasp is then mapped to one of three predefined Barrett hand grasps. Depending on the type of robot grasp, a precomputed grasp strategy is selected. The strategy is further parameterized by the orientation of the hand relative to the object. Experiments in simulated and real environment show the convenience of this method for learning by demonstration purposes.**

I. Introduction

Programming service robots for new tasks puts significant requirements on the programming interface and the user. It has been argued that the Programming by Demonstration (PbD) paradigm offers a great opportunity to unexperienced users for integrating complex tasks in the robotic system [2]. The aim of a PbD system is to use natural ways of human-robot interaction where the robots can be programmed for new tasks by simply observing human performing the task. However, representing, detecting and understanding human activities has been proven difficult and has been investigated closely during the past several years in the field of robotics [3], [4], [5], [6], [7], [8], [9].

In the past, we have studied different types of object manipulation tasks where grasp recognition represents one of the major building blocks of the system [2]. Grasp recognition was performed using magnetic trackers [8], an invasive measurement device.

After the grasp is recognized, two ways of mapping the grasp to a robot were proposed: one based on predefined grasps and another based on an ANN. Although magnetic trackers and datagloves deliver exact values of hand joints, it is desirable from a usability point of view that the user demonstrates tasks to the robot as naturally as possible; the use of gloves or other types of sensors may prevent a natural grasp. This motivates the use of systems with visual input.

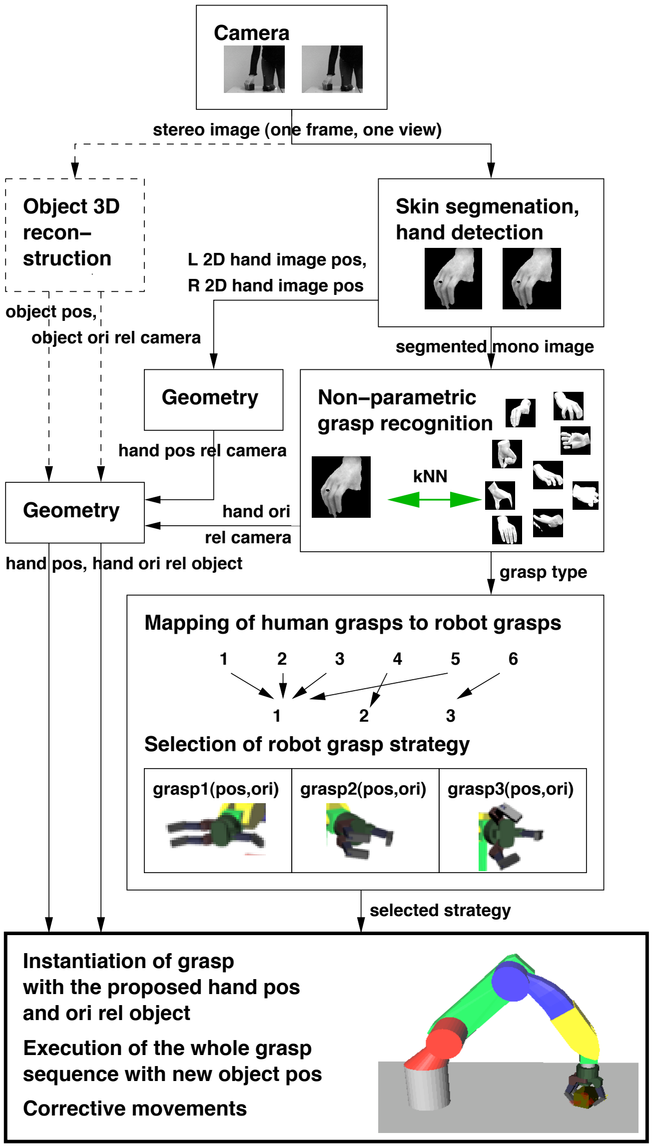
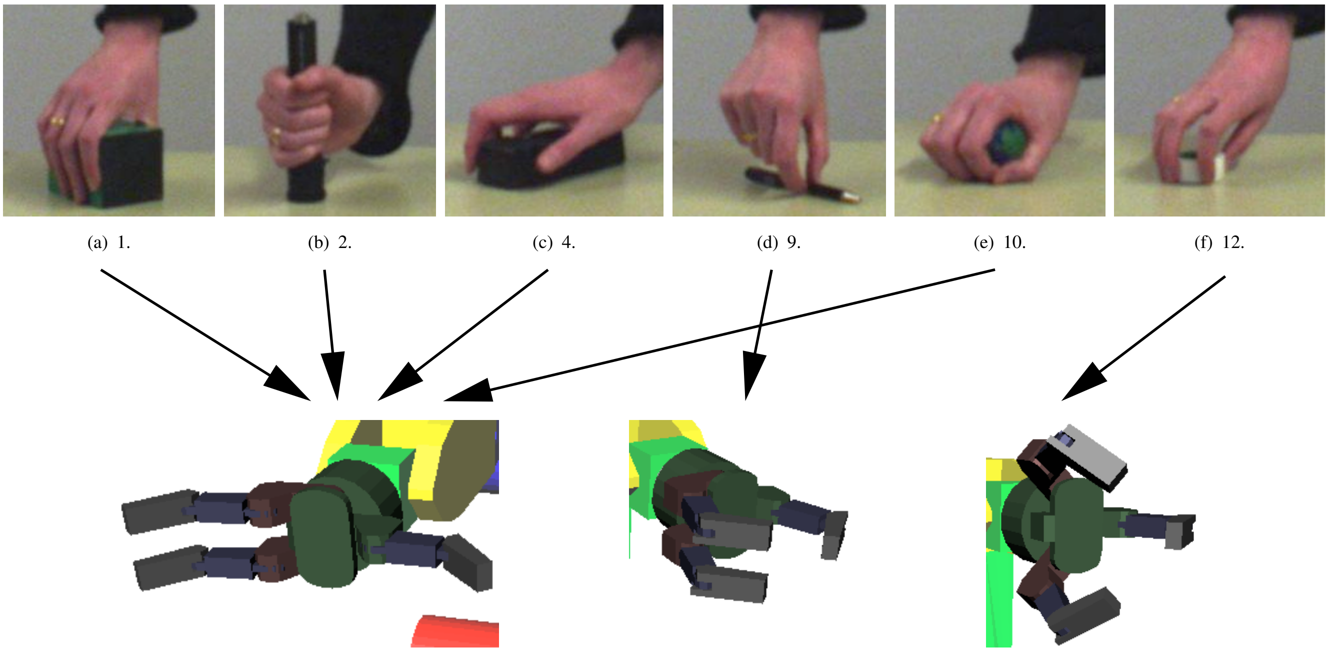


Fig. 1. Human grasps are recognized and mapped to a robot. From one video frame, the hand is localized and segmented from the background. The hand orientation relative to the camera, and type of grasp is recognized by a non-parametric classifier/regressor. The human grasp class is mapped to a corresponding robot grasp, and a predefined grasp strategy, *the whole approach-grasp-retreat sequence*, for that grasp is selected. The strategy is parameterized with the orientation and position of the hand relative to the object. Using this strategy, the robot plans and carries out the grasping action. The focus of this paper is on the grasp execution; details on the grasp recognition are found in [1]. (In our experiments, the object position and orientation were obtained manually.)



(a) 1. (b) 2. (c) 4. (d) 9. (e) 10. (f) 12.

(g) Barrett Wrap. (h) Barrett Two-finger Thumb. (i) Barrett Precision Disc.

Fig. 2. The six grasps (numbered according to Cutkosky's grasp taxonomy [10]) considered in the classification, and the three grasps for a Barrett hand, with human-robot class mappings ((a,b,c,e)→(g), (d)→(h), (f)→(i)) shown. a) Large Diameter grasp, 1. b) Small Diameter grasp, 2. c) Abducted Thumb grasp, 4. d) Pinch grasp, 9. e) Power Sphere grasp, 10. f) Precision Disc grasp, 12. g) Barrett Wrap. h) Barrett Two-finger Thumb, i) Barrett Precision Disc.

Figure 1 outlines the whole mapping procedure. The system consists of three main parts: The human grasp classification, the extraction of hand position relative to the grasped object (with object detection not implemented for our experiments), and finally the compilation of a robot grasp strategy. This paper is mainly focused on the execution of the robot grasp, based upon the instantiated grasping strategy.

An underlying idea in this work is the generation of a robot grasp strategy without tracking the human hand over time or estimating its detailed 3D pose. While articulate 3D reconstruction of the hand is straightforward when using magnetic data or markers, 3D reconstruction of an unmarked hand from images is an extremely difficult problem due to the large occlusion [11], [12], [13], [14], actually more difficult than the grasp recognition problem itself as discussed in Section II. Our method can classify grasps and find their orientation, from a single image, from any viewpoint, without building an explicit representation of the hand, similarly to [13], [15]. Other grasp recognition methods (Section II) consider only a single viewpoint or employ an invasive sensing device such as datagloves, optical markers for motion capture, or magnetic sensors.

The general idea to recognize the human grasp and select a precomputed grasping strategy differs from the traditional way to go about the mapping problem [8]; to recover the whole 3D pose of the human hand, track it through the grasp, and then map the motion to the robot arm. A recognition-based approach such as ours avoid the difficult 3D reconstruction problem, and is also much more computationally efficient since it only requires processing of a single video frame.

The grasp recognition problem is here formalized as the problem of classifying a hand shape as one of six grasp classes, labeled according to Cutkosky's grasp taxonomy [10]. The classes are, as shown in Figure 2a-f, Large Diameter grasp, Small Diameter grasp, Abducted Thumb grasp, Pinch grasp, Power Sphere grasp and Precision Disc grasp. In Figure 2g-i we see the correspondent classes of Barrett hand grasps; due to the lower number of degrees of freedom and fingers in this hand, 4 of the human grasps are mapped to one unique grasp. We should mention here that this distinction between the grasps is not just in terms of preshape of the hand, but also about different strategies for approaching the objects; as pointed by [16] the approach strategy for a precision grasp is different to the approach strategy for a power grasp.

Our paper at IROS 2008 [1] describes in more detail the non-parametric human grasp classification and hand orientation regression. For convenience, a brief description of the grasp recognition method is included in Section III.

In Section IV we explain how to use the data extracted from the database to perform the robot grasp. The type of grasp recognized determines the preshape of the hand in the grasping moment, while the orientation of the hand, together with the object orientation and position, define the approach path. For grasp types where the palm is in contact with the object during the grasp, tactile feedback is also employed for corrective movements during execution of the grasp.

Section V describes a qualitative evaluation of the system, both in a simulated environment and in a real robotic environment.

II. Related Work

A. Grasp planning

In the robotics literature, many solutions to grasp planning have been proposed. Systems for grasp planning can be divided into two main kinds: systems based on the observation of the object to be grasped and systems based solely on the observation of a human performing the grasp. The first kind of solutions rely on the assumption that the appearance of the object in an image can be used either to recognize the object and therefore apply some predefined grasp to it [17], [18], or generate a grasp based on the features of the object [19], [20], [21], [22], [23]. Both [17] and [18] follow the same approach: they create offline a database of optimal grasps for objects whose 3D models are known; then they recognize the objects online as one of the objects in the database and finally they apply on of the optimal grasps. In [19] they extract good grasping points based on object features, and then they perform grasps with a gripper. Different criteria based on 2D geometry of planar objects was used in [23] to retrieve a set of stable grasps. In [20] a box decomposition generated from an object point cloud was used to provide a set of possible grasps. Provided the 3D model of an object, [22] extract from a database which hand pose would fit the shape of the object, based on the normal to the object and hand in different random points. Finally, [24] classify objects based on their affordances (categories like "sidewall-graspable"), so the classification itself determines how to grasp the object. Once they are trained, these systems do not need any external help to perform the grasp. However, it can be really difficult to guess how to grasp an object just based on the appearance. For example, the way to grasp a hammer is not the most natural or most stable for this object, but it is the best for the purpose a hammer is used.

The systems based on a human demonstration can overcome these difficulties, since the teacher shows how the grasp should be exactly performed. However, they are not fully autonomous when they face an object completely new. The complexity of the imitation of a human arm-hand system is high due to the high number of degrees of freedom, the self occlusions and the adaptation to robotic embodiment. We can divide the execution of a grasp into arm movement and hand movement. There are several systems performing imitation of arm [25], [26] or, more generally, upper-body [27], [28], [29]. The arm/upper-body imitation has not so many problems in terms of self occlusion as the hand has. However, it is usually covered by clothes, so it has not a well-defined color and loose clothes can heavily occlude the movements. One approach for grasp imitation would be to track the arm and apply simple predefined grasps (in [30] something similar is performed, but keeping track also of two fingers). However, it has sense to focus our attention in the hand, since perception in biological motion focuses overt attention at the end-effector ([31]). Hand imitation fixes the position of the wrist, end-effector of the arm, so the arm movements are very constrained.

One main point of the hand imitation is how to extract from the human demonstration information that is useful for the purpose of imitation. There are two main branches for this question: model-based approaches try to use regression with some features extracted from the demonstration [11], [32]; appearance-based approaches perform a classification of the hand pose (information useful for the imitation can be extracted from the database). In the first approach we find systems that, based on high level features (like fingertips position), they imitate the whole hand posture [11] or perform a simple mapping to a gripper based on two fingertips [32].

B. Grasp Recognition

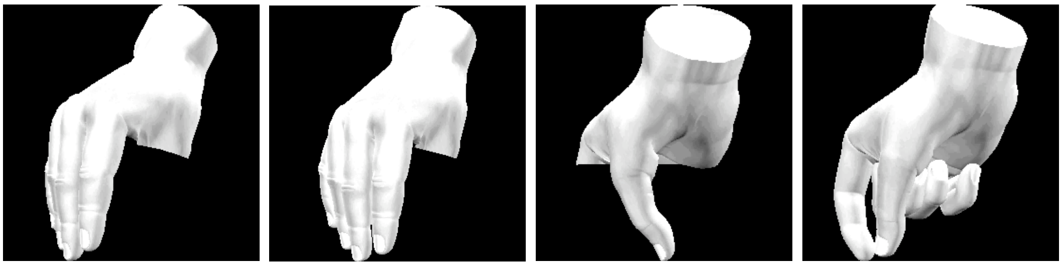
Classification of hand pose is most often used for gesture recognition, e.g. sign language recognition [13], [33]. These applications are often characterized by low or no occlusion of the hands from other objects, and a well defined and visually disparate set of hand poses; in the sign language case they are designed to be easily separable to simplify fast communication. Our problem of grasp recognition differs from this application in two ways. Firstly, the grasped object is usually occluding large parts of the grasping hand. We address this by including expected occlusion in our dataset; occluding objects are present in all example views [1]. Secondly, the different grasping poses are in some cases very similar, and there is also a large intra-class variation, which makes the classification problem more difficult.

Related approaches to grasp recognition [34], [35] first reconstruct the hand in 3D, from infrared images [35] or from an optical motion capture system which gives 3D marker positions [34]. Features from the 3D pose are then used for classification. The work of Ogawara et al. [35] views the grasp recognition problem as a problem of shape reconstruction. This makes their results hard to compare to ours. In addition, they also use a wide baseline stereo system with infrared cameras, which makes their approach difficult to adopt in a case of a humanoid platform.

The more recent work of Chang et al. [34] learns a non-redundant representation of pose from all 3D marker positions – a subset of features – using linear regression and supervised selection combined. In contrast, we use a completely non-parametric approach where the classification problem is transformed into a problem of fast approximate nearest neighbor search [1]. While a linear approach is sufficient in the 3D marker space of Chang et al. [34] , the classes in the orientation histogram space are less Gaussian shaped and more intertwined, which necessitates a non-linear or non-parametric classifier as ours.

Using 3D motion capture data as input, Chang et al. [34] reached an astonishing recognition rate of up to 91.5%. For the future application of teaching of service robots it is however not realistic to expect that the teacher will be able or willing to wear markers to provide the suitable input for the recognition system. 3D reconstructions, although with lower accuracy, can also be achieved from unmarked video

(a) Hand image $\mathbf{H}$. (b) $1, 0.5039, (0, -90, -132)$. (c) $1, 0.5238, (0, -96, -138)$. (d) $1, 0.5308, (0, -96, -132)$. (e) $1, 0.5517, (0, -90, -126)$.



(f) $1, 0.5523, (0, -96, -144)$. (g) $1, 0.5584, (0, -102, -132)$. (h) $1, 0.5870, (0, -90, +108)$. (i) $4, 0.6068, (0, -90, +120)$.

Fig. 3. Distance-weighted nearest neighbor classification. a) Hand view $\mathbf{H}$. b-i) Some of the approximate nearest neighbors to $\mathbf{H}$, with associated grasp class $y_{i,j}$, distance in state-space $d_{i,j}$, and 3D orientation $\mathbf{o}_{i,j}$.

[36], [37]. However, Chang et al. [34] note that the full 3D reconstruction is not needed to recognize grasp type. Grasp recognition from images is thus an easier problem than 3D hand pose reconstruction from images, since fewer parameters need to be extracted from the input. We conclude that the full 3D reconstruction is an unnecessary (and error prone) step in the chain from video input to grasp type.

Our previous work [8] considered an HMM framework for recognition of grasping sequences using magnetic trackers. Here, we are interested in evaluating a method that can perform grasp classification based on a single image only, but it should be noted that the method can easily be extended for use in a temporal framework.

Once the information from the demonstration is extracted, it should be used to perform the robotic mapping. This step is highly dependent on the complexity of the robotic hand: for example, mapping to a gripper has been done just by setting the gripper positions to the position of two fingertips [32]. However, the mapping to more complex hands is usually much more complicated. In [38] the concept of "virtual finger" is introduced: one or more real fingers acting in unison. Kang and Ikeuchi [16] use this concept as an intermediate step in the mapping procedure.

Ekvall [2] presents two different methods for mapping. The first one is based on an ANN whose input is the position of certain parts of the hand (index finger, little finger, palm and chest) and the output is the degrees of freedom of the robotic hand. The second one is based on a database of human hand poses tagged with a predefined mapping to the robotic hand; similar to the approach used here. However, in our approach this database also provide the orientation of the hand in the grasp, while in Ekvall system the orientation is extracted through simulation of several approach vectors.

III. Visual Recognition of Grasp

The content of this section is described in more detail in our IROS 2008 paper [1].

Since the robot grasp strategies are predefined, and only parameterized by the hand orientation, position and type of grasp, there is no need for the human to show the whole grasp procedure; only one time instance is enough (for example, the image that is grabbed when the human tells the robot "now I am grasping").

The input to the recognition method is thus a single monocular image $\mathbf{I}$ from the a camera mounted on the robot. Before fed into the recognition, the image is preprocessed in that the grasping hand is extracted from the background using skin color segmentation.

The segmented image $\hat{\mathbf{H}}$ is cropped around the hand and converted from RGB to grayscale. An example of the resulting hand image $\mathbf{H}$ is shown in Figure 3a.

The classification method is non-parametric; grasp classification and hand orientation regression is formulated as a problem of finding the hand poses most similar to $\mathbf{H}$ in a large database. For the database, a very large set of examples, from many different views, is needed for each grasp.

As it is virtually intractable to generate such training sets using real images, we use a commercial software, Poser 7, to generate synthetic views $\mathbf{H}^{\text{synth}}$ of different hand configurations. 900 views of each configuration were generated, with viewing angles covering a half-sphere in steps of 6 degrees in camera elevation and azimuth; these are the views which can be expected by a robot with cameras above human waist-height. The synthetic hand was grasping an object, whose shape was selected to be typical of that grasp [10]. The object was black (as the background), and occluded parts of the hand as it would in the corresponding real view of

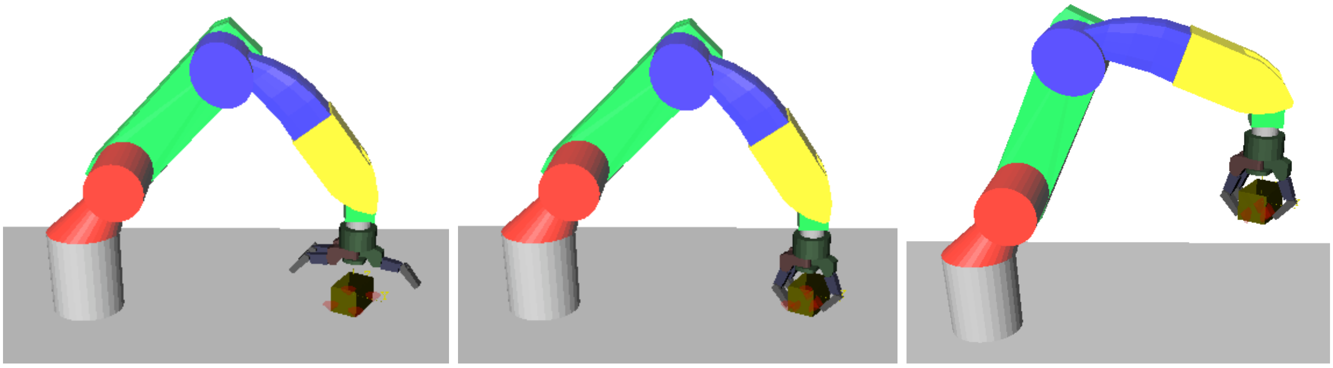


Fig. 4. Barrett Wrap grasp, carried out on the same type and size of object as the human Large Diameter grasp shown in Figure 3a.

that grasp. This will make the synthetic views as similar as possible to the real views (e.g. Figure 3a), complete with expected occlusion for that view and grasp. Figure 3b shows such a database example (the estimated nearest neighbor of Figure 3a).

Each database sample $\mathbf{H}_{i,j}^{\text{synth}}$, where i denotes grasp type and j denotes sample number, has associated with it a class label $y_{i,j} = i$ and a hand-vs-camera orientation $\mathbf{o}_{i,j} = [\phi_j, \theta_j, \psi_j]$, i.e. the Euler angles from the camera coordinate system to a hand-centered coordinate system.

To find the grasp class $\hat{y}$ and orientation $\hat{\mathbf{o}}$ of an unknown grasp view $\mathbf{H}$ acquired by the robot, a distance-weighted k-nearest neighbor (kNN) classification/regression procedure is used. First the set of k nearest neighbors to $\mathbf{H}$ in terms of Euclidean distance between gradient orientation histograms obtained from the grasp images [1] are retrieved from the database.

As an exact kNN search would put serious limitations on the size of the database, an approximate kNN search method, Locality Sensitive Hashing (LSH) [39], is employed. LSH is a method for efficient approximate nearest neighbor search. The computational complexity of retrieval is $\mathcal{O}(DN^{\frac{1}{1+\epsilon}})$ where ϵ is the approximation factor. This gives sublinear performance for any $\epsilon > 0$. Figure 3b-i show approximate nearest neighbors to the hand in Figure 3a.

From the found approximate nearest neighbors, the estimated class of $\mathbf{H}$ is found as a distance-weighted selection of the most common class label among the k nearest neighbors, and the estimated orientation as a distance-weighted mean of the orientations of those samples among the k nearest neighbors for which $y_{i,j} = \hat{y}$. (The cyclic properties of the angles is also taken into account in the computation of the mean.) As we can see in Figure 3i, the orientation of a sample from a different class has very low correlation with the real orientation, simply because the hand in a different grasp has a different shape. Therefore, only estimates with the same class label as $\hat{y}$ are used in the orientation regression. All in all, the dependency between the hand view space and the global Euler angle space is highly complex, and that is why it is modeled non-parametrically.

IV. Example-Based Mapping of Grasp to Robot

The estimated grasp class as well as hand and object orientation and position are used to instantiate a robotic grasp strategy, as illustrated in Figure 1.

A human-to-robot grasp mapping scheme is defined depending on the type of robot hand used; in our experiments we use a Barrett hand with three types of grasps as shown in Figure 2. The type of robot grasp defines the preshape of the robot hand.

The hand orientation estimate $\hat{\mathbf{o}}$ relative to the camera, along with the hand position estimate and the estimated position and orientation of the grasped object relative to the camera, are used to derive the estimated position and orientation of the human hand relative to the object, as depicted in Figure 1. The estimation of object position and orientation is assumed perfect; this part of the system is not implemented, instead the ground truth is given in the simulations.

In contrary to related grasp approaches [40], the robot here does not explore a range of approach vectors, but instead directly imitates the human approach vector, encoded in the hand position and orientation relative to the object. This leads to a much shorter computational time at the expense of the non-optimality of the grasp in terms of grasp quality. However, since the the selection of robotic preshape has been guided, the stability of the robotic grasp will be similar to the human one, leading to a non-optimal but successful grasp provided that the errors in the orientation and position estimate are sufficiently small.

Based on the estimated type of grasp, the system first differentiates between volar and non-volar grasp ([16]), i.e., whether there is a contact between the palm and object or not. Volar grasps are the Large Diameter, Small Diameter, Abducted Thumb and Power Sphere grasps (see Figure 2). The contact between the palm and the object makes it possible to use tactile feedback to perform corrective movements during the final part of the grasp execution. This makes the grasping less sensitive to visual errors. (Corrective movements can also be guided by vision, but this functionality is not yet included in our system.)

The volar grasping is performed in the following order: First, the robot adopts the hand orientation and preshape corresponding to the estimated human grasp. The robot hand then approach the object until it detects contact in the palm sensor. After that, it closes the hand, and retreats. Two different ways of approaching the object are used, based on the orientation of the hand; if the human fingers are parallel or close to parallel to the table plane the object is approached from the side, otherwise it is approached from the top.

Looking closer at the four volar grasps, the Abducted Thumb grasp differs somewhat from the others; while the rest of them fit well the Barrett Wrap preshape, this one does not. The Barrett hand cannot grasp two faces of a thin object while keeping the contact between the palm and the object. For this reason the Abducted Thumb grasp is better performed as a Barrett Two-finger Thumb grasp.

The non-volar grasps, which have no contact between the palm and the object, are the Pinch grasp and the Precision Disc grasps (see Figure 2). Since there is no contact between the tactile sensor in the palm and the object, the grasp is in our system performed "blindly", without any feedback. For this reason, the non-volar grasps depend heavily on the precision of the of the object position and orientation estimation. The Pinch and Precision Disc grasps differ only in the preshape; otherwise the robot grasp strategy is identical: The robot adopts the hand orientation and preshape corresponding to the estimated human grasp. The robot hand then approach the object until it is at a predefined distance over the object. After that, it closes the hand, and retreats.

V. Experimental Results

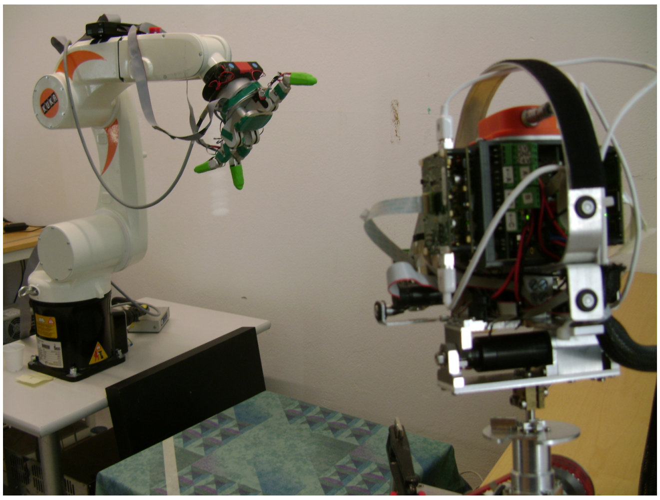
To illustrate how the grasp classification can be employed for human-to-robot mapping in a pick-and-place scenario, both a simulated robot arm and a real robot arm are controlled with parameterized pre-defined grasping strategies as illustrated in Figure 1.

A. *Simulated grasping with GraspIt*

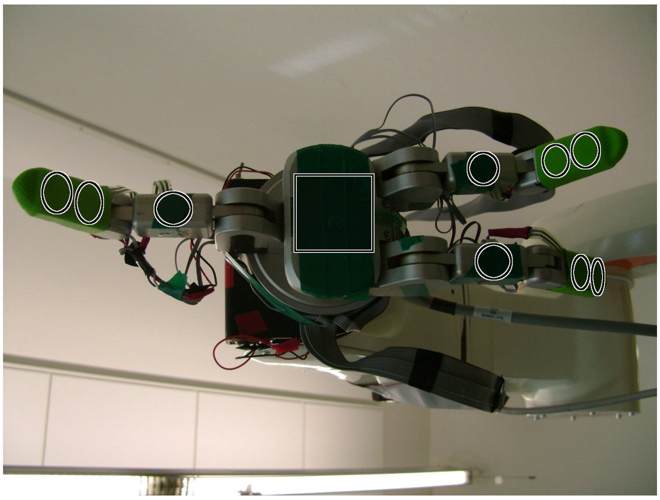
We first test the grasp strategy instantiation in the simulated environment GraspIt. This makes it possible to control the errors in estimated relative hand-object orientation and position.

An analysis of the robustness to position errors can be found in [40]. For an optimally chosen preshape, there is a error window ≥ 4 cm $\times$ 4 cm about the position of the object, within which the grasps are successful. The positioning of the robot hand can also be improved by fusing the estimated human hand position with an automatic selection of grasping point based on object shape recognition [19].

The robustness to orientation errors depends greatly on the type of grasp and object shape. We investigated the robustness of the Barrett Wrap grasp with an approach vector perpendicular to the table (Figure 4). We get good results for orientation errors around the vertical axis of up to 15 degrees. As a comparison, the mean regression error of this orientation [1] is on the same order as the error window size, 10.5 degrees, which indicates that the orientation estimation from the grasp classifier should be used as an initial value for a corrective movement procedure using e.g. the force sensors on the hand.



(a) Head Observing Robot



(b) Hand

Fig. 5. Grasping environment. a) A robotic stereo head observing the table where the experiments are performed, and a KUKA arm with the Barrett hand mounted. b) The Barret hand with the sensors marked.

B. *Real grasping with a KUKA arm*

The grasp mapping is then integrated with the grasp recognition [1], and implemented on a robot with a KUKA arm and a Barrett hand. The environment used is shown in Figure 5.

The robotic head observes the scenario while the object is grasped by the human, and captures an image when the ENTER key is pressed on the robot's keyboard. We plan to replace this by a speech recognition system. The image is passed to the grasp recognition [1], which returns the type of grasp and the position and orientation of the hand. With those parameters, the policy is selected and performed with the KUKA arm and Barrett hand showed in Figure 5. Tactile feedback is given by the pressure sensors in Figure 5b.

The scenario, illumination and subject is different to the experiments in [1], but we get similar results in the classification. Large diameter, small diameter and abducted thumb are correctly classified most of the time, while pinch grasp, power sphere and precision disc grasp are sometimes confused with the power grasp. In terms of orientation, the typical error is around 10 to 15 degrees, which is acceptable in the execution of the grasp, as discussed above.

The object position is given manually, with an error of ± 3 cm. The position error did not inflict on the grasp execution, except when performing Precision Disc grasp with a ball, which rolled when the hand was not centered over the ball.

Figure 6 shows the robot being shown four different grasps (Large Diameter, Abducted Thumb, Pinch and Precision Disc, respectively), mapping them and performing the corresponding grasp (Barrett Wrap, Barrett Two-finger Thumb, Barrett Two-finger Thumb and Barrett Precision Disc, respectively).

VI. Conclusions

In this paper, a method for classification of grasps, based on a single image input, was presented. A grasping hand was represented as a gradient orientation histogram; a 2D image-based representation. A new hand image could be classified as one of six grasps by a kNN search among large set of synthetically generated hand images.

The grasp classification retrieved the grasp class together with the orientation of the hand. This information was used to select and parameterize a policy for performing the grasp on a robot in a simulated and a real environment. The experiments indicated a need for sensor feedback during the execution of the grasp on the robot.

A. Future Work

The system presented here can be improved in several ways. The database of hand poses should include more objects of different sizes, and more advanced non-parametric regression methods could be employed for estimating grasp type and hand pose.

Moreover, the addition of visual servoing would improve considerably the performance in grasps without tactile feedback (volar grasps) or grasps when the tactile feedback fail due to the limitations of the hand sensors.

Finally, we will investigate the performance using more humanoid-like robot hands, which are expected to improve the imitation performance.

VII. Acknowledgments

This research has been supported by the EU through the project PACO-PLUS, FP6-2004-IST-4-27657, and by the Swedish Foundation for Strategic Research.

References

[1] H. Kjellström, J. Romero, and D. Kragić, "Visual recognition of grasps for human-to-robot mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[2] S. Ekvall, "Robot task learning from human demonstration," Ph.D. dissertation, KTH, Stockholm, Sweden, 2007.

[3] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 6, pp. 799–822, 1994.

[4] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends in Cognitive Sciences*, vol. 3, no. 6, pp. 233–242, 1999.

[5] A. Billard, "Imitation: A review," in *Handbook of Brain Theory and Neural Networks*, M. Arbib, Ed., 2002, pp. 566–569.

[6] K. Ogawara, S. Iba, H. Kimura, and K. Ikeuchi, "Recognition of human task by attention point analysis," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000, pp. 2121–2126.

[7] M. C. Lopes and J. S. Victor, "Visual transformations in gesture imitation: What you see is what you do," in *IEEE International Conference on Robotics and Automation*, 2003, pp. 2375–2381.

[8] S. Ekvall and D. Kragić, "Grasp recognition for programming by demonstration tasks," in *IEEE International Conference on Robotics and Automation*, 2005, pp. 748–753.

[9] S. Calinon, A. Billard, and F. Guenter, "Discriminative and adaptative imitation in uni-manual and bi-manual tasks," in *Robotics and Autonomous Systems*, vol. 54, 2005.

[10] M. Cutkosky, "On grasp choice, grasp models and the design of hands for manufacturing tasks," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, 1989.

[11] J. Rehg and T. Kanade, "Visual tracking of high dof articulated structures: An application to human hand tracking," in *European Conference on Computer Vision*, vol. 2, 1994, pp. 35–46.

[12] E. Ueda, Y. Matsumoto, M. Imai, and T. Ogasawara, "A hand-pose estimation for vision-based human interfaces," in *IEEE Transactions on Industrial Electronics*, vol. 50(4), 2003, pp. 676–684.

[13] V. Athitsos and S. Sclaroff, "Estimating 3D hand pose from a cluttered image," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. 432–439.

[14] C. Schwarz and N. Lobo, "Segment-based hand pose estimation," in *Canadian Conf. on Computer and Robot Vision*, 2005, pp. 42–49.

[15] H. Murase and S. Nayar, "Visual learning and recognition of 3-D objects from appearance," *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[16] S. B. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception: Mapping human grasps to manipulator grasps," *IEEE Transactions on Robotics and Automation*, vol. 13, no. 1, pp. 81–95, 1997.

[17] A. Morales, P. Azad, T. Asfour, D. Kraft, S. Knoop, R. Dillman, A. Kargov, C. Pylatiuk, and S. Schulz, "An anthropomorphic grasping approach for an assistant humanoid robot," in *International Symposium of Robotics*, 2006.

[18] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *IEEE International Conference on Robotics and Automation*, 2003, pp. 1824–1829.

[19] A. Saxena, J. Driemeyer, J. Kearns, and A. Y. Ng, "Robotic grasping of novel objects," in *Neural Information Processing Systems*, 2006.

[20] K. Hübner and D. Kragić, "Selection of robot pre-grasps using box-based shape approximation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[21] C. Borst, M. Fischer, and G. Hirzinger, "A fast and robust grasp planner for arbitrary 3D objects," in *IEEE International Conference on Robotics and Automation*, 1999, pp. 1890–1896.

[22] Y. Li and N. Pollard, "A shape matching algorithm for synthesizing humanlike enveloping grasps," in *IEEE International Conference on Humanoid Robots*, 2005.

[23] A. Morales, P. J. Sanz, and A. P. Pobil, "Vision-based computation three-finger grasps on unknown planar objects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[24] M. Stark, P. Lies, M. Zillich, J. Wyatt, and B. Schiele, "Functional object class detection based on learned affordance cues," in *Computer Vision Systems*, 2008.

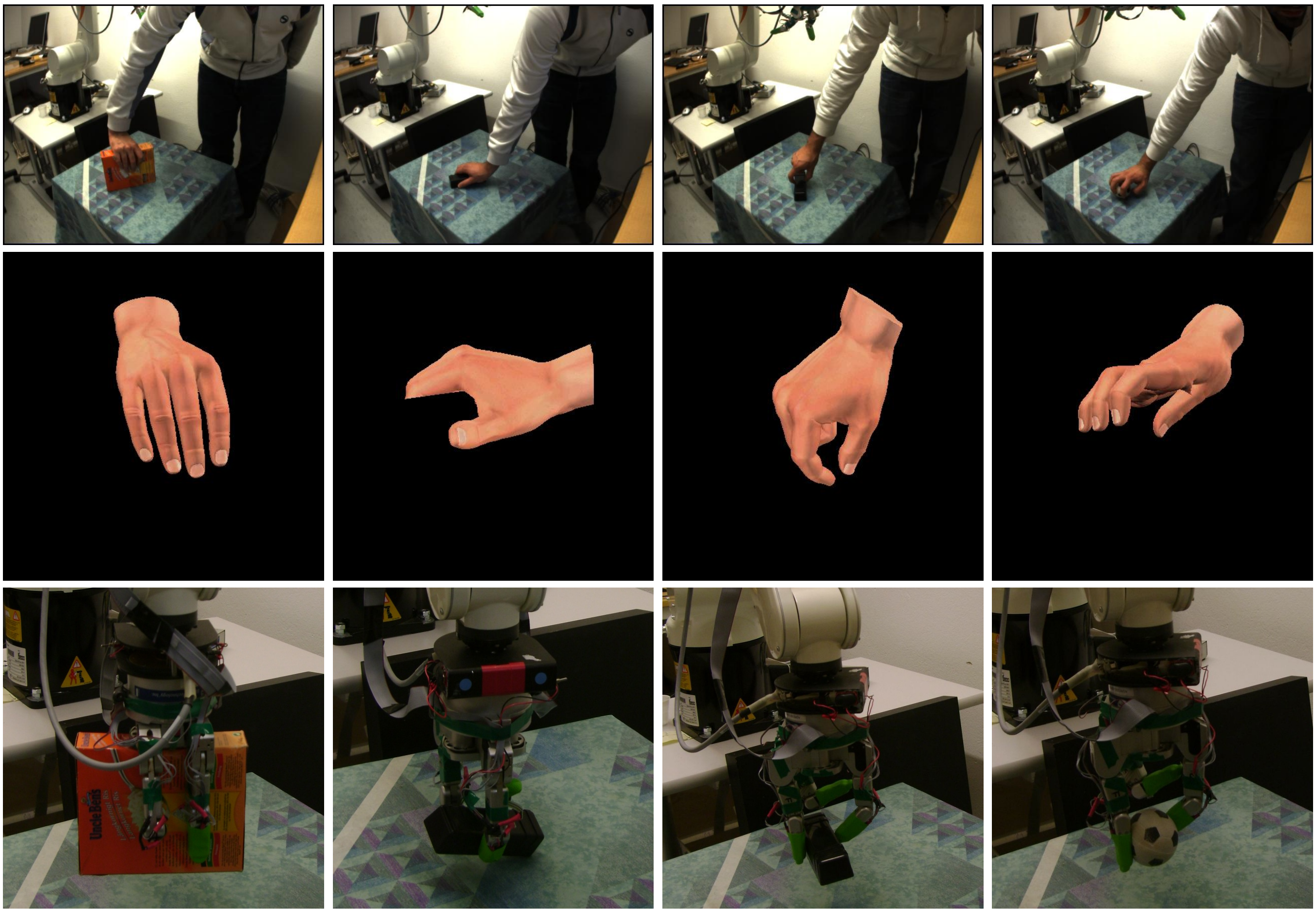
[25] A. Billard, S. Calinon, and F. Guenter, "Discriminative and adaptive imitation in uni-manual and bi-manual tasks," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 370–384, 2006.

[26] A. Billard, Y. Epars, S. Schaal, and G. Cheng, "Discovering imitation strategies through categorization of multi-dimensional data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 2003, pp. 2398–2403.

[27] D. Ormoneit, H. Sidenbladh, M. J. Black, and T. Hastie, "Learning and tracking cyclic human motion," in *Advances in Neural Information Processing Systems 13*, 2001, pp. 894–900.

[28] M. Riley, A. Ude, K. Wade, and C. G. Atkeson, "Enabling real-time full-body imitation: a natural way of m-ansferring human movement to humanoids," in *IEEE International Conference on Robotics and Automation*, 2003, pp. 2368–2374.

[29] P. Azad, A. Ude, T. Asfour, and R. Dillmann, "Stereo-based markerless



(a) 1 → Barrett Wrap. (b) 4 → Barrett Two-finger Thumb. (c) 9 → Barrett Two-finger Thumb. (d) 12 → Barrett Precision Disc.

Fig. 6. Execution of grasps in a real robot environment. First row shows images grabbed by the robotic head at the grasping moment, second row shows the nearest neighbors to the first row pictures in the database, and the third row shows the robot execution of the same grasp. a) Large Diameter grasp,1, mapped to Barrett Wrap. b) Abducted Thumb grasp, 4, mapped to Barrett Two-finger Thumb. c) Pinch grasp, 9, mapped to Barrett Two-finger Thumb. d) Precision Disc grasp, 12, mapped to Barrett Precision Disc.

human motion capture for humanoid robot systems," in *IEEE International Conference on Robotics and Automation*, 2007, pp. 3951–3956.

[30] D. Herzog, V. Krüger, and D. Grest, "Exemplar-based parametric hidden markov models for recognition and synthesis of movements," in *Vision, Modeling, and Visualization*, 2007.

[31] M. J. Matarić, "Sensory-motor primitives as a basis for imitation: Linking perception to action and biology to robotics," in *Imitation in Animals and Artifacts*, K. Dautenhahn and C. Nehaniv, Eds., 2000.

[32] J. Triesch, J. Wieghardt, E. Mael, and C. Malsburg, "Towards imitation learning of grasping movements by an autonomous robot," in *International Gesture Workshop*, 1999, pp. 73–84.

[33] Y. Wu and T. S. Huang, "Vision-based gesture recognition: A review," in *International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, 1999, pp. 103–115.

[34] L. Y. Chang, N. S. Pollard, T. M. Mitchell, and E. P. Xing, "Feature selection for grasp recognition from optical markers," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.

[35] K. Ogawara, J. Takamatsu, K. Hashimoto, and K. Ikeuchi, "Grasp recognition using a 3D articulated model and infrared images," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, 2003, pp. 1590–1595.

[36] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical bayesian filter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1372–1384, 2006.

[37] E. Sudderth, M. I. Mandel, W. T. Freeman, and A. S. Willsky, "Visual hand tracking using non-parametric belief propagation," in *IEEE Workshop on Generative Model Based Vision*, 2004.

[38] M. A. Arbib, T. Iberall, and D. M. Lyons, "Coordinated control programs for movements of the hand," in *Hand function and the neocortex. Experimental Brain Research Supplemental 10*, A. W. Goodwin and I. Darian-Smith, Eds., 1985.

[39] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *International Conference on Very Large Databases*, 1999, pp. 518–529.

[40] J. Tegin, S. Ekvall, D. Kragić, B. Iliev, and J. Wikander, "Demonstration based learning and control for automatic grasping," in *International Conference on Advanced Robotics*, 2007.