Proceedings of the ICAR 2009
14th International Conference on Advanced Robotics
Munich, Germany, June 22 - 26, 2009

Extraction of grasp-related features by human dual-hand object exploration

Krzysztof Charusta, Dimitar Dimitrov, Achim J. Lilienthal, Boyko Iliev
Center of Applied Autonomous Systems (AASS)
Department of Technology, Örebro University, Sweden

***Abstract*— We consider the problem of objects exploration for grasping purposes, specifically in cases where vision based methods are not applicable. A novel dual-hand object exploration method is proposed that takes benefits from a human demonstration to enrich knowledge about an object. The user handles an object freely using both hands, without restricting the object pose. A set of grasp-related features obtained during exploration is demonstrated and utilized to generate grasp oriented bounding boxes that are basis for pre-grasp hypothesis. We believe that such exploration done in a natural and user friendly way creates important link between an operator intention and a robot action.**

I. INTRODUCTION

Mastering robot grasping capabilities have been always considered of major importance since this skill allows to execute other more complicated tasks. Thus, the robot ability to perform stable grasps is a fundamental. There were many approaches that addressed problem of robot grasping but each of theme suffer from some drawbacks that our solution tries to solve.

In the group of approaches that the 3D model of an object is known various methods based on friction cones [1] or form- and force-closure criteria can be applied to perform stable grasps. However, mostly a priori knowledge is not available or it lacks informations about object properties like surface texture or mass center.

A lot of work has been done in the area of visual object recognition and modeling for grasping. For example, [2] presents a learning algorithm for grasps prediction for parallel gripper based on 2D images. Other authors [3], [4] try to estimate most stable grasping points for multi-fingered hands based on different vision systems and grasp stability criteria. In [5], [6] a method for pre-grasp selection is proposed by decomposing object shape into minimum bounding boxes. However, the robustness of vision based methods to object recognition and localization is compromised if visual clues are absent or in cluttered environment. In addition it is not possible to obtain the information about the object mass or its surface friction solely based on vision based recognition, without interaction with the environment.

Some of the mentioned vision problems are solvable using active exploration. Interactive robot perception has been studied in [7] where a robot manipulates the environment to obtain properties of an object (kinematic model). However, such autonomous exploration is still far from human exploration capabilities. Thus, a human demonstration based approach has clear advantages.

A Programming by Demonstration (PbD) approach (see [8] for a recent overview) has been popular especially in area of programming grasping and manipulation tasks [9], [10], [11], [12]. A variety of human tracking devices have been used, following [13], the most popular one due to its robust and accurate hand pose acquisition is a data glove. A double glove experimental setup has been used by [14] for programming of dual-arm manipulation tasks. In [15] the posture of a bare human hand is tracked for grasp acquisition using a camera. Additional hardware is not necessary but the vision system requires careful presentation planning to make the hand and the object visible.

In the light of the adduced work, to overcome the drawback of vision based methods and utilize the benefits of human demonstration, we propose a novel approach to object modelling, by dual-hand human tactile exploration. In this way the knowledge about the object is extended by a set of features that captures the human way of object handling.

In this work we let the user present an object freely using both hands, without restricting the object pose, with only hardware constraints that imply that all grasps must be fingertip grasps. The proposed exploration method takes advantage of the fact that a human operator usually unconsciously simplifies the grasping task by selecting one of only a few different prehensile postures appropriate for a given object and task. In this way the problem searching space is limited to a set of most plausible grasping approaches.

The humans act differently depending on object shape and handling task [16]. Proposed exploration method is not directed on emphasizing the geometry of an object where the operator would rather intentionally focus on showing significant features of the target like edges, corners, roundness. This can be done much better using vision systems and 3D scanning. Our method is focused on capturing features important for object handling in pick and place scenarios that are hardly available using vision. Moreover, it is a useful extension for all vision based method and mixing them together should provide very robust grasping oriented recognition system.

The remainder of this paper is structured as follows. In Section II our experimental setup, the employed sensor devices and the experimental scenario it is presented. Section III details the proposed exploration algorithm. Section IV

introduces a clustering algorithm for detection of graspable region and grasp related features extraction method. Finally, Section V presents experimental results and an evaluation of the accuracy of the object pose. Section VI concludes the paper with a summary and a discussion of future work.

II. Experimental setup

To register a human demonstration robustly and efficiently proper experimental setup is needed. The system should fulfill two main requirements, firstly, acquire of the configurations of two hands in real-time, secondly, it should allow the user to explore an object in natural and convenient way. For this reason in our setup a vision based *PhaseSpace motion capture* system [17] for human hands tracking has been used.

The advantages of the system are high accuracy and sampling rate, robustness to changing light conditions and, in contrast to 2D- and 3D vision methods, it allows exploration of objects without assumptions of any specific visual clues. Of course, it suffers from occlusion as any other vision based system.

A. Sensor Gloves

To let the user explore an object fast and freely two gloves are used. Each glove has nine diodes and five force sensing resistors (FSR) used as tactile sensors emplace on the glove as depicted on Fig. 1.

Fig. 1. Motion capture gloves with diodes and FSR tactile sensors emplacement.

Each diode blinks in a unique pattern that can be identify and its 3D position is tracked using a set of five stereo cameras placed around the working area (see Fig. 2). The camera positions are arranged so they cover most of the working space.

As a glove enhancement, we attached passive tactile sensors (FSR) to the glove fingertips which, as presented in [18], are good enough to detect grasp actions. Since the tactile sensors are placed only on the fingertips only precision grasps can be registered. Some remarks about the sensor use have to be made. Despite quite large sensor surface, the material rigidness and sensor emplacement allows only prismatic like grasps. The tactile sensors are rather slippery so the operator has to grasp firmly during all experiments and rather choose rough surfaces. This fact has also an advantage because obtain grasping regions are characterized by good grip.

The data from both the motion capture system and the tactile sensors is visualized so the operator can easily see online how the object is represented and what is the algorithm's interpretation of human demonstrations. Though motion capture system allows to capture data with frequency of $480Hz$ our system capability due to accuracy improvements and visualization rendering is reduced to $25Hz$.

B. Assumption

Several assumptions about the demonstration and obtained information has to be made.

- A1 The demonstration is aimed at presenting grasping possibilities for handling tasks.
- A2 The human hand, while holding an object is considered as a rigid body. As performed experiments have shown the human hand (excluding wrist) tends to act as a rigid body when performing grasping tasks. Thus the wrist diode is not considered as a part of the hand configuration.
- A3 No vision system for object tracking has been used.
- A4 All grasps performed by human are stable, static, precision grasps according to force closure criterion.
- A5 A grasp is collected if at least the thumb and three other fingers are in contact with an object.
- A6 No assumption neither about an object size or geometry, nor about gripper that will be used are made.

C. Experiment scenario

An Operator picks an object of interest from the table and handles it freely using both hands. The collected point cloud appears on the screen so a visual feedback is possible. Resultant data is then clustered and grasp-related features are associated with the object. The two main steps of the method are briefly described below.

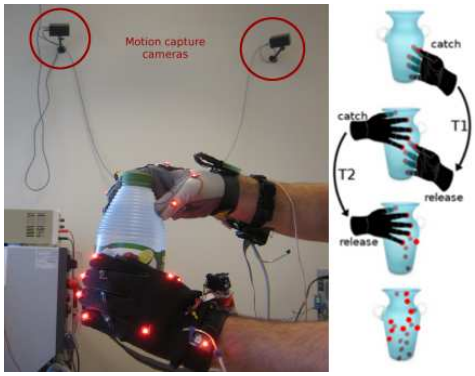

Fig. 2. On the left object exploration using two gloves with two motion capture cameras in the background. On the right schematic object exploration sequence with collected points marked red.

1) Object exploration: The whole exploration consists of a sequence of fingertip grasps registered by the tactile sensors and the motion capture system. During the manipulation *atom sets* of points, that each represents separate, fingertip

grasp, are collected. Because the user is allowed to change the holding hand, a transformation between grasps is calculated to keep all points in a one object coordinate frame. This calculation is possible because a human hand holding an object is treated as a rigid body. To reinforce the rigid body assumption and improve accuracy of transformations averaging techniques, presented in later sections, have been used. Object exploration is described in details in III.

2) Features extraction: Resulting data from the object exploration step is a collection of atom point sets that together create 3D point cloud. The point cloud is sparse, so only rough deduction about the object geometry is possible, however it is good enough to show graspable regions on the object surface. Moreover, a human way of approaching the object is captured in collected data. Atom sets are firstly clustered into separate graspable regions - *bodies*. Secondly, for every grasp an *approach vector* and associated with it *grasp oriented bounding box* (GOBB), that bounds all points that belong to the same body, are generated. The method for generation of approach vectors and GOBBs is described in Section IV. As a result, the object is approximated with a set of overlapping GOBB with respective approach vectors. Bounding box has been chosen as an approximation primitive, however, such fitting can also be done with other primitives like cylinders or spheres.

III. Object Exploration

A. Point cloud building

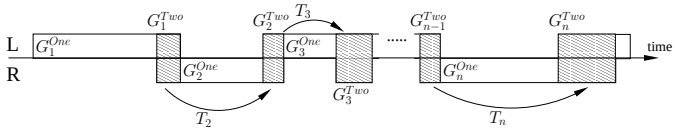

Fig. 3. Time line with periods of one-hand- and dual-hand-grasps. Transformations necessary to keep all points in one object coordinate frame.

During the exploration a 3D point cloud is obtained that represents an approximate geometry of the object. The object can be lifted from the table and the operator is allowed to translate and rotate it freely using both hands, thus, an algorithm that keeps all collected points in one object coordinate frame is necessary. As Fig. 3 depicts the exploration contains a sequence of grasps alternately using left or right hand, with intermediate states in between when both hands hold the object. The goal is to collect grasping points and calculate a sequence of transformations that models object move.

Let us describe this exploration as a sequence of overlapping states

$$\mathcal{G} = \{G_1^{One}, G_2^{One}, \ldots, G_i^{One}, \ldots, G_n^{One}\}, \tag{1}$$

where G_i^{One} indicates the $i-th$ consecutive time period when one-hand grasp is performed on the object. Additionally, two-hand grasp state

$$G_i^{Two} = G_i^{One} \wedge G_{i+1}^{One}, \tag{2}$$

is defined for cases when both hands holds the object. Let us define $\mathbf{l} = [x, y, z]^T$ as a diode coordinate in global reference frame. The hand configuration in the state G_i^{One} is described as a set of 9 diodes coordinates

$$h_i = \{\mathbf{l}_j | j = 1, \ldots, 9\}, \tag{3}$$

where each position $\mathbf{l}_j$ corresponds to diode j on the glove. Let us denote $\hat{h}_i$ as an average hand configuration which will be basis for further calculations in this section. Method of obtaining $\hat{h}_i$ based on a few h_i configuration is described later in section III-B.

Through the whole time period when only one hand is in contact (state G^{One}) neither hand configuration, relative to the object, nor contact points change. Thus, it is sufficient to collect only one hand configuration $\hat{h}_i$ for every state G_i^{One}. It is reasonable to collect this hand configuration at the state G_i^{Two} when it overlaps with G_i^{One}, because two hand grip increase rigidness of the grasp and improve precision of further transformation calculations.

The transformation matrix T_i between every two hand configurations $\hat{h}_{i-1}$ and $\hat{h}_i$ from consequent states G_{i-1}^{Two} and G_i^{Two}, has to be calculated. Since coordinates of diodes in global reference frame and their correspondence between hand configurations are known our problem becomes to determine unknown transformation T_i of the Least Squares (LS) problem

$$\hat{h'}_i \approx T_i \hat{h'}_{i-1}, \tag{4}$$

where $\hat{h'}_i$ and $\hat{h'}_{i-1}$ are selections of those diodes that are visible in both configurations $\hat{h}_i$ and $\hat{h}_{i-1}$. An algorithm that solves this problem using singular value decomposition (SVD), as described in [19], has been used.

Point cloud building can be seen as an iterative process, where in every state G_i^{Two} an atom point cloud, that contains positions of diodes of fingers that are in contact with an object and are visible, is saved. Atom point cloud is denoted as

$$A_i = \{\mathbf{f}, \ldots, \mathbf{f}_r\}, \quad A_i \subset \hat{h_i}, \quad 3 < r \leq 5, \tag{5}$$

where $\mathbf{f}$ is a diode position of a visible finger in contact, r is a number of such fingers. All previously collected points have to be transformed into the actual frame. Let us finally denote P_i as a set that contains all points: previously collected and transformed to the actual state i and newly saved atom point cloud.

$$P_i = \{P'_{i-1}, A_i\}, \tag{6}$$

where

$$P'_{i-1} = T_i \cdot P_{i-1}, \tag{7}$$

B. Accuracy improvements

There are two main sources of errors in transformation calculations that should be eliminated. Firstly, Least Square fitting works well only if a diode position, relative to other diodes, doesn't change between hand configurations $\hat{h}_{i-1}$ and $\hat{h}_i$. Secondly, calculations might be impossible if number of diodes visible in both configurations is smaller than three.

It has been noticed during experiments that even when the grasp is firm and stable, the hand makes moves, imperceptible for a human, but significant enough to make

transformation calculations inaccurate. Thus, wrist diode has been discarded in transformation calculation, because its position changes significantly. For the rest of the diodes, averaging algorithms presented below were applied.

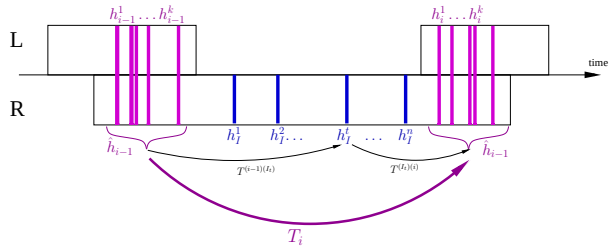

Fig. 4. Transformation T_i in detail. Example selection of hand configurations for averaging process and inter hand configuration depicted.

1) Key hand configuration averaging: In the time period S_i^{Two} k hand configurations with the greatest number of visible diodes are selected and an average hand configuration $\hat{h}_i$ is calculated on them (see Fig. 4)

$$\hat{h}_i = \frac{1}{k}\sum_{j}^{k} T^{jk} h_i^j, \tag{8}$$

where T^{jk} is transformation matrix, calculated using mentioned LS fitting, between hand configuration h^j and h^k.

2) Inter hand configurations: Another implemented accuracy improvement is to find m intermediate hand configurations h_I between states G_{i-1}^{Two} and G_i^{Two} and calculate transformation T_i as a mean of transformations that go through these intermediate hand configurations

$$T_i = \frac{1}{m}\sum_{t=1}^{n} T^{(I_t)(i)} T^{(i-1)(I_t)}, \tag{9}$$

where $T^{(I_l)(i)}$ is a transformation matrix between hand h_i and h_I^t. LS fitting algorithm has been used also in this case.

3) Gloves calibration: Diodes that the motion capture system tracks are mounted on the finger nails on the outer side of the hand. This means that all tracked positions are measured with an offset between a diode center and a real finger contact point. It is removed by applying calibration strategy in which operator touches a table surface with previously know plane equation. Normal to plane distance between diodes positions and the table surface is calculated. Obtained offset values allow to estimate contact points with an object f'_j as depicted on Fig. 5.

IV. Features extraction

In this section a method for extracting grasp-related features is presented. It utilizes point cloud collected with the method presented in Section III. Firstly, distinguishable graspable regions of an object are detected - *bodies*, secondly, human-like approach vectors are found, finally *grasp oriented bounding boxes* are generated. All those features are associated with the explored object.

A. Points clustering

In order to distinguish between different graspable parts of the object and to segment collected contact points clusterization of atom point clouds is performed. To do that parameters on which clusterization is based have to be calculated.

For every atom point cloud A_i that contains thumb position a *grasp direction* D_i is defined as the first principle component vector of a set of middle points

$$M = \{m_j | j = 1 \dots r-1\}, \tag{10}$$

where m_j is a middle point between the thumb and a finger j (see Fig.5). A *grasp center* c_i is defined as a mean value

$$c_i = \frac{1}{r-1}\sum_{j=1}^{r-1} m_j. \tag{11}$$

Then all atom point clouds are segmented using unsupervised nearest neighbor clustering algorithm in two steps. Firstly, based on angle between grasp directions measured as an absolute value of cosine between two grasp direction vectors D_a and D_b

$$d = \left| \frac{D_a \cdot D_b}{\|D_a\|\|D_b\|} \right|, \tag{12}$$

and secondly, based on Cartesian distance between grasp centers. Threshold values have been chosen empirically to $20\,\mathrm{deg}$ and $80mm$ respectively. As a result, the point cloud P is clustered into separate segments - *bodies*, that each will be later approximated using primitive shape.

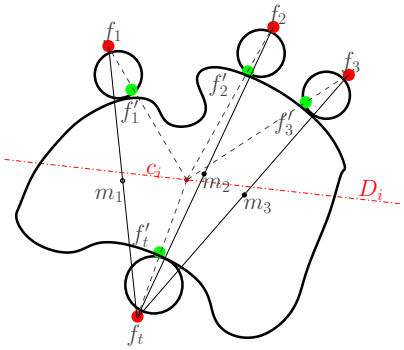

Fig. 5. An example of one-hand grasp with 4 fingers visible projected onto the plane. f - diodes positions marked red; f' - estimated contact points after calibration applied marked green; D_i - grasp direction, c_i - grasp center.

B. Approach vectors

One of the advantages derived from a human exploration is a set of human-like approach vectors that is assigned to every body segment obtained in clusterization process. As depicted on Fig. 6 every single grasp, as long as proper diodes have been seen, has assigned approach vector which has its initial point a_i calculated as

$$\mathbf{a}_i = \frac{\frac{\mathbf{l}_{h1}+\mathbf{l}_{h2}}{2} + \mathbf{l}_w}{2}, \tag{13}$$

where $\mathbf{l}_{h1}$, $\mathbf{l}_{h2}$ are mid-hand diodes and $\mathbf{l}_w$ denote a wrist point. The approach vector is oriented towards grasp center c_i.

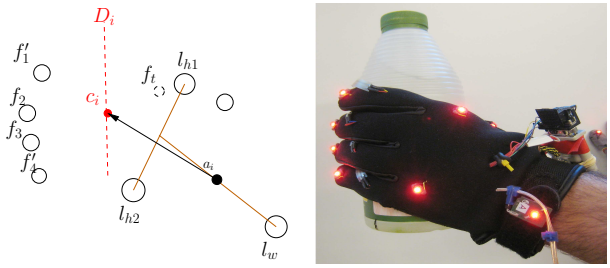

Fig. 6. Schematic single grasp and corresponding real hand configuration. Hand direction (D_i), hand center point (c_i), finger contact points (f') and approach vector depicted.

C. Grasp oriented bounding boxes

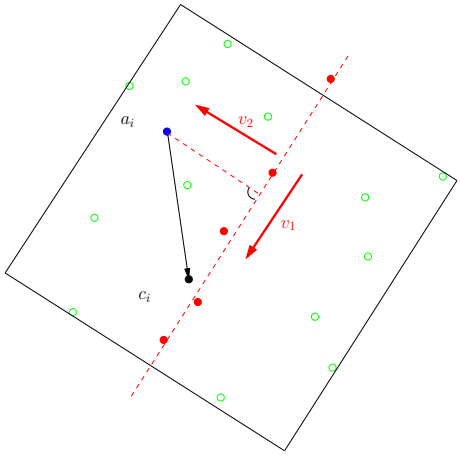

Fig. 7. An example of grasp oriented bounding box in 2D. Principle vectors v_1, v_2 marked red.

For every performed grasp, that has the approach vector and the grasp center, an GOBB is generated that bounds all contact points of a given body (see Fig. 7). To obtain its orientation a set of orthogonal vectors is calculated. The first principle vector is common for all grasps from a given body. It is obtained by taking the first principle component of a set of grasp centers from given body. The second principle vector is defined as a perpendicular to the first one, going through the initial point a_i of the approach vector. The third vector is orthogonal to previous two. As a result for every segmented body multiple GOBBs are generated. It represents approach choices natural for a human that might be also considered as a pre-grasp states in grasping tasks.

V. EXPERIMENTS

A set of experiments has been performed to check performance of described methods of exploration and grasp-related features extraction. Variety of common objects like: bottle, hammer, drill, phone receiver, orange or book have been chosen to demonstrate how the method behaves for different shapes. Selection of the results of performed experiments is depicted in Fig. 8. Proposed algorithm for features extraction works well for different types of objects and gives a suggestion about how an object should be grasped. A few comments about performed experiments can be found below.

GOBB decomposition is not view-dependent as in vision based methods but it is grasp-dependent. Thanks to free manipulation, generated bounding boxes envelope also these parts that might be invisible in visual exploration. At the same time, they bound only those parts that have been touched. Fig. 8(f) shows that the operator explored mainly the hammer shaft, since only this part has rubber surface that is easy to hold, especially using our gloves with slippery tactile sensors.

As presented in Fig. 8(d) and Fig. 8(e) segmentation divides objects into separated bodies that describe roughly object shape. We can distinguish two bodies of the phone receiver. It is even more noticeable in the case of a drill where the object is also divided into two bodies that correspond to drill core and handle. Notice that both bodies have complete different approach vectors characteristic. The vectors for drill handle part are very consistent and suggest only one way of grasping this part. In the same time, approach vectors for the cylindrical core of the drill are much more spread and no strict approach is suggested.

Obtained features like approach vectors, size, rough shape should be considered as basis for pre-grasp generation. Besides, some clues about object intrinsic properties like surface texture, graspable regions are acquired. Although they are just clues, they can be useful in planning of grasping and manipulation tasks.

A. Object pose error

To evaluate the accuracy of calculated transformations, an error of object pose in global coordinate frame has been checked. To do that a cylindrical bottle has been explored for at least ten grasps and then placed on the table in the origin of the global coordinate system. In series of ten experiments biggest position error was not greater than $0.5cm$ and orientation error not greater than $7\,\mathrm{deg}$.

VI. CONCLUSION AND FUTURE WORKS

This paper presents a novel, dual hand object exploration method. It is oriented towards grasp-related features extraction such like human-like approach vectors, size, rough shape which can be basis for pre-grasp calculation. The presented method for two hands exploration is fast since only one demonstration is enough to point out key features of an object. It is also a human intuitive which is be especially important for the end user. Accuracy tests have shown that the method is reliable enough to find explored object after the demonstration and can be used even without any vision system. However, as pointed out at the beginning, integrating of both together will make this system complete and both part will supplement each other. A question regard utilization of grippers with size and grasp model different from human hand stays open and should be considered as a research continuation. Further work include grasp experiments using an anthropomorphic artificial hand e.g. the KTHand [20]

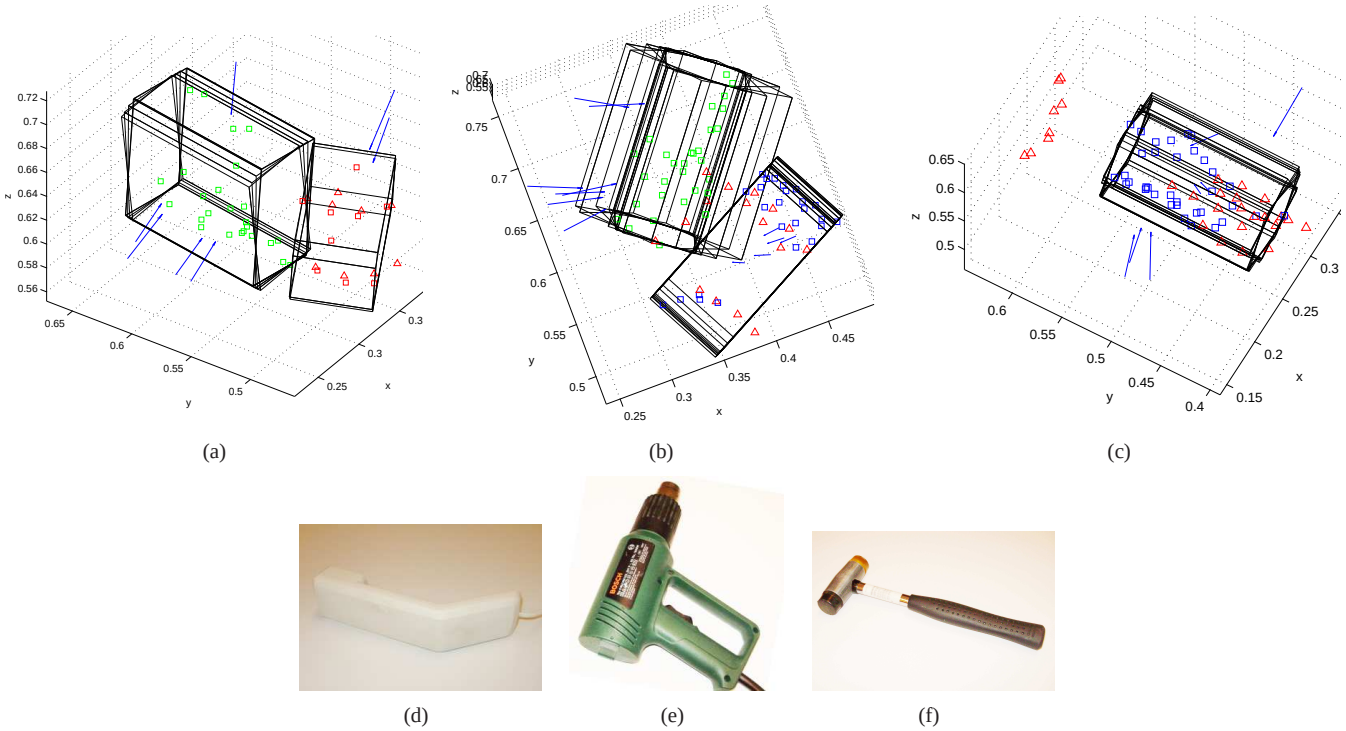

Fig. 8. Sample experiment results with corresponding real objects pictures. Grasp oriented bounding boxes (black lines) and approach vectors (blue vectors) presented. Square - clustered contact points and triangle - unclustered contact points. (a),(d) - phone, (b),(e) = drill, (c), (f) - hammer

and integrating the presented system with a visual object recognition system.

REFERENCES

[1] M. Mason and J. K. Salisbury, *Robot Hands and the Mechanics of Manipulation*. Cambridge, MA: MIT Press, 1985.

[2] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *Int. J. Rob. Res.*, vol. 27, no. 2, pp. 157–173, 2008.

[3] T. Yoshikawa, M. Koeda, and H. Fujimoto, "Shape recognition and grasping by robotic hands with soft fingers and omnidirectional camera," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, May 2008, pp. 299–304.

[4] A. Morales, P. J. Sanza, Angel, and A. H. Fagg, "Vision-based three-finger grasp synthesis constrained by hand geometry," *Robotics and Autonomous Systems*, vol. 54, no. 6, pp. 496–512, 2006.

[5] K. Huebner and D. Kragic, "Selection of robot pre-grasps using box-based shape approximation," *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 1765–1770, Sept. 2008.

[6] K. Huebner, S. Ruthotto, and D. Kragic, "Minimum volume bounding box decomposition for shape approximation in robot grasping," *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 1628–1633, May 2008.

[7] D. Katz and O. Brock, "Manipulating articulated objects with interactive perception," *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 272–277, May 2008.

[8] A. Billard and R. Siegwart, "Robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 47, no. 2-3, pp. 65–67, 2004.

[9] J. J. Steil, F. Röthling, R. Haschke, and H. J. Ritter, "Situated robot learning for multi-modal instruction and imitation of grasping," *Robotics and Autonomous Systems*, vol. Special Is, no. 47, pp. 129–141, 2004.

[10] R. Dillmann, "Teaching and learning of robot tasks via observation of human performance," *Robotics and Autonomous Systems*, vol. 47, no. 2-3, pp. 109–116, 2004.

[11] Castellini, Claudio, Orabona, Francesco, Metta, Giorgio, Sandini, and Giulio, "Internal models of reaching and grasping," *Advanced Robotics*, vol. 21, no. 13, pp. 1545–1564, 2007.

[12] R. Zollner and R. Dillmann, "Using multiple probabilistic hypothesis for programming one and two hand manipulation by demonstration," *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 3, pp. 2926–2931 vol.3, Oct. 2003.

[13] D. J. Sturman and D. Zeltzer, "A survey of glove-based input," *IEEE Computer Graphics and Application*, 1994.

[14] R. Zollner, T. Asfour, and R. Dillmann, "Programming by demonstration: dual-arm manipulation tasks for humanoid robots," in *Intelligent Robots and Systems, 2004. IROS 2004. Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 1, Sept.-2 Oct. 2004, pp. 479–484 vol.1.

[15] M. Hueser, T. Baier, and J. Zhang, "Learning of demonstrated grasping skills by stereoscopic tracking of human head configuration," *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 2795–2800, May 2006.

[16] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 269–279, 1989. [Online]. Available: http://dx.doi.org/10.1109/70.34763

[17] "Phasespace motion capture." [Online]. Available: http://www.phasespace.com/

[18] R. Zollner, O. Rogalla, and R. Dillmann, "Integration of tactile sensors in a programming by demonstration system," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 3, 2001, pp. 2578–2583 vol.3.

[19] W. Gander, "Least squares fit of point clouds," *Solving problems in scientific computing using Maple and MATLAB (3rd ed.)*, pp. 339–349, 1998.

[20] J. Tegin, J. Wikander, and B. Iliev, "A sub 1000 euro robot hand for grasping – design, simulation and evaluation," in *Int. Conf. on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, Coimbra, Portugal, Sept. 2008.