Grasping Movements Recognition in 3D Space using a Bayesian Approach

Diego R. Faria, Hadi Aliakbarpour and Jorge Dias

Abstract — **In this work, grasping movements recognition in 3D space is presented. We also present the idea of a database implementation with different sensors data in grasping and handling tasks scenarios for our future works. Multi-sensor information for grasp tasks require sensors calibration and synchronized data with timestamp that we started to develop to share with the researches of this area. In the scenario presented in this work we are performing the grasp recognition combining 2 different types of features from reach-to-grasp movements. By observing the movements from different subjects we can perform a learning phase based on histogram techniques using the segmented data. By applying Bayes rule by continuous classification based on multiplicative updates of beliefs we can classify the movements. We developed an automated system to estimate and recognize two types of reach-to-grasp movements (e.g., for top and side grasps). These reported steps are important to understand some human behaviors before the object manipulation and can be used to endow a robot with autonomous capabilities, like showing how to reach some object for manipulation.**

I. Introduction

Grasping movements have been the focus of interest of many researches, including areas like neuroscience and robotics. Studies in neuroscience field, human reach-to-grasp trajectories are analyzed to verify the brain areas that are activated with determined tasks. Investigation about human trajectories is useful to analyze the hand shape, pose and velocity, i.e., the kinematic changes to the reach-to-grasp movement in people with Parkinson disease or post-stroke. It is useful to verify the performance and behaviors of these people concerning movement stability, motor coordination, etc. In robotics field, hand trajectories are useful for human-robot interaction using gestures to interact with social robots and also for complex tasks like imitation learning. In this work we want to show the estimation and recognition of grasp movements by a Bayesian approach. Analyzing these movements we can be able to understand some human behaviors during the hand journey to reach and grasp an object. This information can be used to endow robots with human-like actions, i.e., using the movements before the object manipulation or object displacement. Beside of reach-to-grasp analysis, this work can be useful also for gesture recognition for human robot interaction. We also intend to make a contribution with database of grasp movements using different sensors and different scenarios for grasping tasks showing some useful sensors calibration for some specific grasp tasks.

This work is partially supported by the European project Handle ICT-23-16-40. Diego Faria and Hadi Aliakbarpour are supported by Portuguese Foundation for Science and Technology (FCT). Diego Faria, Hadi Aliakbarpour and Jorge Dias are with Institute of Systems and Robotics – University of Coimbra – Polo II, 3030-290 Coimbra, Portugal (diego, hadi and jorge)@isr.uc.pt.

II. Related Work

Bayesian models are used in [1] to classify gestures from images sequences. Tracking of human hands and face are used based on skin-color features. The proposed application is towards human-robot interaction. The human actions are interpreted and mapped to the robot actions. They have contributed also with Laban Movement Analysis that helps to identify useful low-level features and to develop a classifier of expressive actions. Images sequence are used in [2] for hand tracking and hand shape representation when a person is gripping a mug. The authors proposed a method for hand shape representation that characterizes the finger-only topology of the hand, using cepstral coefficients. Techniques of speech signal processing are used for that. This work shows hand shape recognition classified as top-grab, side-grab, flat-hand and handle-grab when the hand is close to object. Some works show human motion tracking which are important for different applications inside robotics field such as learning human motion models for recognition in vision and learning primitives from motion capture [3]-[6].Several works concerning grasping involves the learning of object affordances in which some of them use different sensors data towards finding different ways to grasp a determined object, such as the work presented by [7].

In our previous work [3], we developed an application to segment trajectories to find 2D changes in direction features like up, down and line for its classification. In that work we have used second order derivative to analyze the evolution of the trajectory finding features analysing just the x and y axes of a 3D trajectory ignoring other features like diagonal, forward and backward directions that could improve the classification. The learning was based on histogram techniques. The classification results were satisfactory but we also obtained undesired results as false negative and recognition of the trajectory with low probability.

In this work, features from 3D trajectories (3D changes in direction) are extracted for grasping movement recognition ugin a probabilistic approach.

III. SENSORS DATABASE: SCENARIOS AND APPLICATIONS

We built a database of different sensors data (Fig.1) in different scenarios of grasping and handling tasks. The main idea is a contribution for this research field with 3D grasp movements acquired from different sensors: 3D trajectories from magnetic tracking system; images sequences from monocular and stereo; fingers flexure during the movement acquired from data-glove; force applied in the object during the manipulation (tactile sensors) and points of interest trough a subject sight (eye tracker).

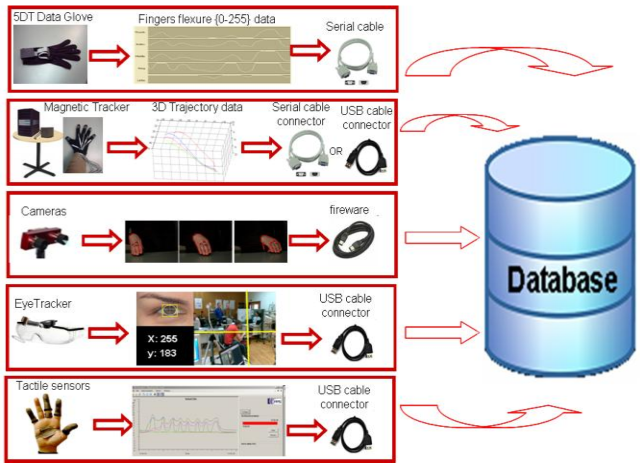


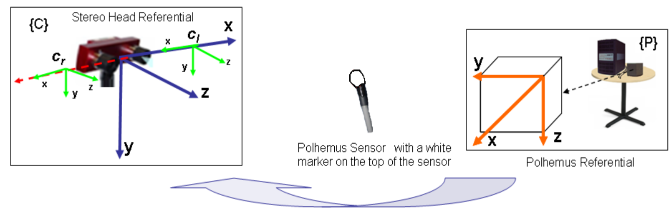
Fig.1. Grasping movement database acquired in different scenarios with multimodal data.

The sensors data stored in the database need to be synchronized by a timestamp during the data collection. Beyond of sensors data, the sensors specifications, calibration parameters and transformations matrices will be stored also in the database. Calibration between the sensors is needed and the transformation results are also stored.

A. *Sensors Calibration Step*

In order to combine both data, from vision and magnetic tracker device, or even to use one as ground-truth, in this work we have done a useful calibration between the Polhemus Liberty 240/8 tracking device [9] and Videre STH-MDCS3-9cm stereo camera [10] to acquire a transformation to re-project the 3D points of the tracker device frame of reference in the image plane and vice-versa. This calibration is also useful for 3D object shape representation using multimodal data, integrating the visual cues from stereo with data from grasp exploration (i.e., using the magnetic tracker attached to the fingertips for contour following, acquiring the object shape by the fingers movements around the object).

The calibration allows us to see a 3D point in the local reference frame of tracker device to the stereo camera reference frame. The first step of this calibration is to acquire the intrinsic and extrinsic parameters of the stereo camera. The Polhemus device give us the 3D points related to its frame of reference, so that we can use the strategy of using a white tape on the sensor and then we can recognize this marker in the image, obtaining the 3D point after the camera calibration (Fig.2). We collected 30 images (from left and right cameras), acquired at same instant of the 3D point from the tracker device sensor in different positions and orientations. The tracker sensor was attached to a tripod on a red piece of paper for easy localization in the image. This idea is originally inspired from auto calibration method between multi-cameras by Svoboda in [11] where a laser pointer to get different views point was used.



$^{\{P\}}T_{\{C\}}$

Fig.2 Calibration strategy: Using a white tape on the sensor we can easily localize the marker in the images to compute the 3D point corresponding to point of the tracker device.

The stereo camera and the tracker reference frames, {C} and {P} respectively, are rigid to each other. Collecting two set of 3D corresponding points in two coordinate references, ${}^c p = \{ {}^c p_i \mid i = 1,\dots,N \}$ and ${}^p p = \{ {}^p p_i \mid i = 1,\dots,N \}$ we compute the following equation to acquire a 3D point from a {P} to {C}:

$$ {}^c p = {}^p R_c \, {}^p p + {}^p t_c \tag{1} $$

To compute ${}^p R_c$ and ${}^p t_c$ (rotation and translation matrices of the homogeneous transformation) Arun's method described in [12] has been used which is based on an algorithm to find the least-squares solution of R and t using singular value decomposition (SVD) of a 3x3 matrix.

Fig.3 shows the result of the calibration: the magnetic tracker sensor with a white tape attached to a tripod and the re-projection of its 3D point is represented as a yellow point in the image.

Tab.1. Average re-projection error (AE) and the standard deviation (SD)

	N= 7	N=10	N=13	N=15
AE	12.363	8.9170	7.3334	6.4914
SD	3.450	3.092	2.923	2.825

Table 1 shows the average re-projection error values, in pixels, according to the number of 3D points used. The average error of the proposed calibration decreases when the method uses a higher number of points. It is possible to consider that for N = 20 points, the calibration method is stable.

Fig.3. Re-projection of the 3D point (yellow color point inside of the circle) of the magnetic tracker in the image plane. The magnetic sensor with a white tape is attached to a tripod on a red piece of paper to easily localize the sensor in the images to compute the 3D points in the camera frame of reference.

Fig.4 and Fig.5 show the evolution of estimates of the rotation and translation matrices acquired from the calibration according to the number of points used.

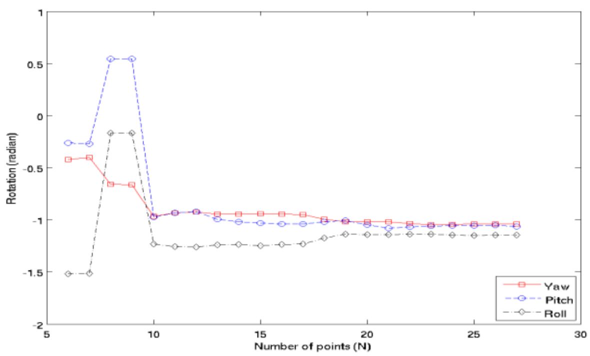


Fig.4. Evolution of the rotation matrix estimate by the calibration method according to the number of points used in the approach.

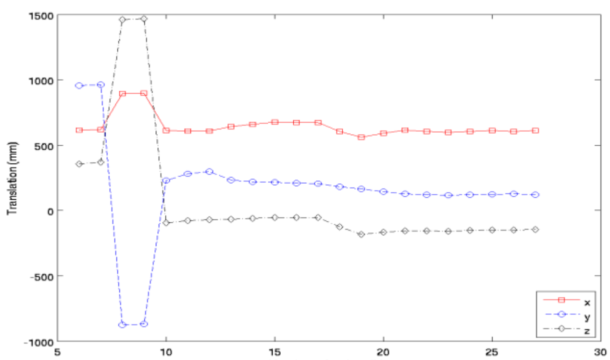


Fig.5. Evolution of the translation matrix estimate by the calibration method according to the number of points used in the approach.

Other sensors require individual calibration like the 5DT data-glove [13] for fingers flexure data. The dynamic range may differ with the persons hand sizes. The calibration by the 5DT software normalizes the effect of different dynamic ranges for different hand sizes. For its calibration the dynamic range is computed as follows:

$$R = V_{\max} - V_{\min} \tag{2}$$

where R is the dynamic range; V_{max} is the maximum output value (flexed hand) and V_{min} is the minimum output value (flat hand). A normalization process is necessary and for that R is used, for example, lets work through the thumb, V_{min} and V_{max} are 40 and 206 respectively , so that $R =$ 166 . To scale the measured values across the full R value (256 values), the normalization is compute as follows:

$$N = (M - V_{\min})\left(\frac{255}{R}\right) \tag{3}$$

where $\boldsymbol{N}$ is the normalized value; M is the measured value, for the thumb.

The calibration parameters and all information correlated are stored in the database for each specific scenario of grasp and manipulation tasks. All sensors used in each specific task have temporal information for synchronization, a timestamp that is also stored with the sensors data.

Through multi-sensors information we can combine cues to better understand how humans achieve a grasping. Some questions can be posed related to how humans perform the grasp (Fig.6).

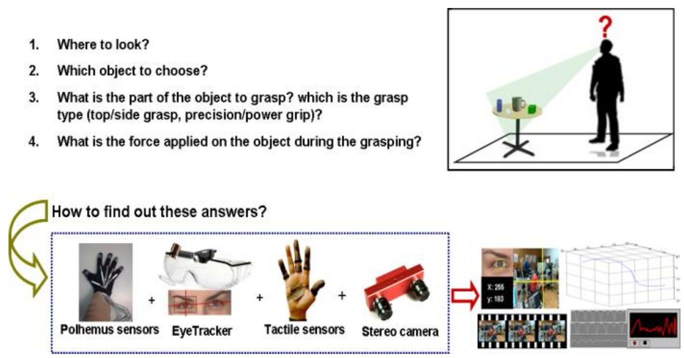


Fig.6. Some questions of how to achieve a grasping and a possible answer to solve this problem trough multi-senor cues.

Multimodality can assist to solve ambiguities and also to map the corresponding knowledge to a robot, for instance: trajectory, force, fingers flexure during a manipulation task. Many human demonstrations using different sensors modalities for a specific task are stored in a dataset and can be used afterwards to learn and replicate the same task by a robot. Other studies can also be done with the stored data, such as the estimation of the object that will be chosen for grasping and the estimation of possible contact points on the object analyzing the person sight (trough eye-tracker device), etc.

IV. Experimental Setup

In this work, we are using the same scenario used in our previous work [8]. We have used Polhemus Liberty tracking device to track hand trajectories performed by humans. We have attached five sensors to a glove for acquiring 3D hand trajectories, allowing us analyzing the fingers behaviors during their journey to the object. Another sensor was attached to object to have a priori knowledge of the object position and the size of the trajectory (e.g. difference between initial hand position and the object sensor). Two reach-to-grasp movements are performed for classification:

Top-Grasp and Side-Grasp (Fig.7). Fig.8 shows our current experimental setup for 3D movements recognition.

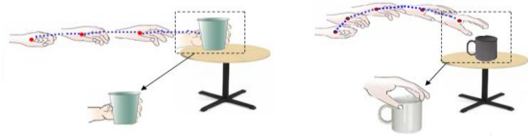


Fig.7. (a) – Side-grasp; (b) Top-grasp.

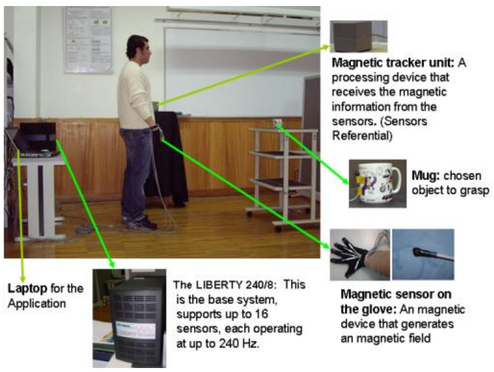


Fig.8. Experimental setup for movement demonstrations and classification.

V. Segmentation And Features Extraction

The segmentation of a movement is empirically done by segmenting the trajectory into specific parts and the features extracted are curvatures and hand orientation in 3D space. The features extraction is based on our previous work [14]. We have used the cylindrical coordinate system (r, θ, h) to detect the features. The possible directions combining h and θ information are: *up*, *down*, *left*, *right*, *up-left*, *up-right*, *down-left*, *down-right and non-movement*. Details about the steps for these features detection are in [14]. Another feature extraction was adopted, the hand orientation along the trajectory by approximating the hand plane using 3 sensors on the fingertips (index, middle and ring finger fingers). Afterwards, two probabilities tables for each trajectory can be built from the learning phase, one for the curvatures detection and another for hand orientation. The segmentation process is performed in each part of a normalized trajectory (i.e., all trajectories are normalized to the same scale). In this work, we have split the trajectory in 8 parts, extracting features in each of these parts. More information about these steps can be found in [14].

VI. Learning and Classification of Grasp Movements

Computational models for human perception and action has been explored by researches. Some studies about human brain reports that Bayesian methods have achieved success in creating computational theories for perception and sensorimotor control [15]. Based on these successful applications of Bayesian theory, in this work we developed our approach using Bayesian techniques as described in the next subsections.

A. Grasp Learned Tables

Given a set of observations to represent a type of Grasp G, at some displacement d, we have the probability of each type of curvature C in each part of a trajectory represented as $P(C \mid G\ D)$. The same is applied for hand orientation, so that we have $P(O \mid G\ D)$ where O represent all possible hand orientation (top or side).
For each movement (demonstrated by a subject) is built a histogram to store the probability distribution of the features. The learned table is a mean histogram calculated from all top grasp and all side grasp demonstrations. For more details see [14]. Fig.9 shows our learned grasp tables for curvatures detection and Fig.10 shows the learned tables for hand orientation detection.

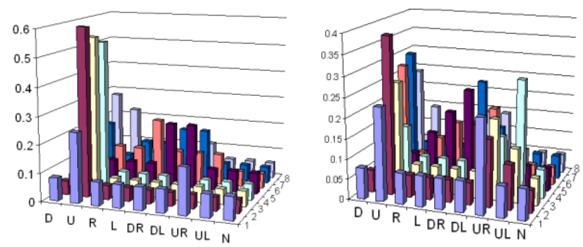


Fig.9. Left image represents the top-grasp curvatures learned table and right image side-grasp learned table.

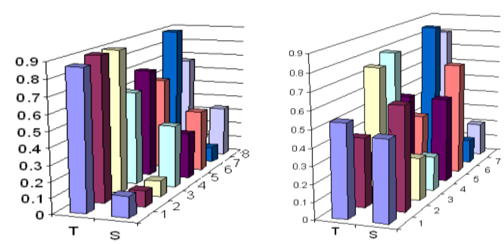


Fig.10. Left image represents the top-grasp hand orientation learned table and right image side-grasp learned table.

B. Classification Model using Bayesian Techniques

Bayesian classification models have already proven their usability in gesture recognition systems [3] [8] [14]. Based on these studies we adopted a Bayesian model for grasp recognition analyzing the reach-to-grasp movements. The estimation and recognition of a type of grasp happens along of a trajectory that is being performed by a subject. In each hand displacement (after a time instant), the probability of each type of grasp is updated. In our previous work [14] we have presented Bayesian models for trajectory classification, one using curvatures features, another using hand orientation features and a third one as an ensemble, combining both posteriori by a mixture model were the weights for each model is obtained by an uncertainty measure. In this work we are simplifying the strategy, showing that a Naïve Bayes classifier as simple a dynamic Bayesian network using the curvatures (changes in direction) and hand orientation features can present good results.

To understand the General Grasp Recognition Model some definitions are done as follows:

1. g is a known grasp from all possible G (Grasp types);
2. c is a certain value of feature C (Curvature types);

3. o is a certain value of feature O (hand orientation types);
4. i is a given index from all possible parts composed of a distance D (1/8 of a trajectory size).

The probability $P(c \mid g\ i)$ that a feature C has certain value c can be defined by learning the probability distribution $P(C \mid G\ D)$. The probability $P(o \mid g\ i)$ that a feature O has certain value o can be defined by learning the probability distribution $P(O \mid G\ D)$. Knowing $P(c \mid G\ i)$, $P(o \mid G\ i)$ and the prior $P(G)$, we are able to apply Bayes rule and compute the probability distribution for G given a hand displacement i concerning the hand displacement of the learned table and the features c and o. Initially, the grasp G is a uniform distribution and during the classification it is updated applying Bayes rule:

$$P(g_{top} \mid c\, o\, i) \propto P(c\, i \mid g_{top})P(o\, i \mid g_{top})P(g_{top}) \tag{4}$$

$$P(g_{side} \mid c\, o\, i) \propto P(c\, i \mid g_{side})P(o\, i \mid g_{side})P(g_{side}) \tag{5}$$

Equations (5) and (6) can be rewritten and represented as follows:

$$P(G \mid C\, O\, D) = \frac{P(C\, D \mid G)\, P(O D \mid G)\, P(G)}{\sum_j P(C\, D \mid G_j)\, P(O D \mid G_j)\, P(G_j)} \tag{6}$$

The main idea here is to use the online classification when someone is performing a trajectory. The posterior probability of a current hand displacement becomes the prior for the next displacement (probabilistic loop). We formulate the equation as recursive way. Assuming that each hand displacement we can find new curvatures and new hand orientation, we can then express the online behaviour by using the index i that represents a certain displacement performed by the person in the reach-to-grasp movement. The rule for classification is based on the higher probability value, being necessary obtaining a certain confidence (e.g. 0.7). We expect that a reach-to-grasp movement that is being performed by a subject to grasp the mug by top or side grasp will produce a grasp hypothesis with a significant probability.

C. *Preliminary Results of Movements Classification*

Fig.11 shows an example of side-grasp trajectory performed by a subject and table 2 shows the answer of our approach along this trajectory classifying it.

Table 3 shows the results of classification for 10 trials performed by different subjects (side-grasp trajectories). A false negative value (trial 5) happened due to the similarity between side and top grasp trajectories. A deep study and tests performing much more trials need to be done with this classification model for further analysis. By now we noticed that the reach-to-grasp movements (top- and side grasp) differs more at the end of the trajectory when the hand has more changes in direction and in its pose (orientation). This happens when the hand is preparing to form a hand configuration for a specific grasp. The object pose and size influences trajectory and type of grasping.

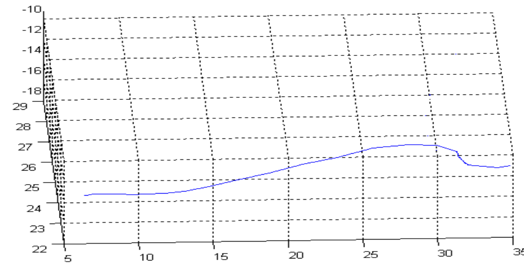


Fig.11. Side-grasp trajectory (raw data measured in inches).

Tab.2. Classification of the trajectory shown in Fig. 11

Trajectory Parts	Top%	Side%
1	0.11	0.88
2	2.44	97.56
3	2.44	97.56
4	0.04	99.96
5	0.04	99.96
6	0.07	99.93
7	0.01	99.99
8	**0.01**	**99.99**

Tab.3. Result of 10 trials of side-grasps. In 10 trials we obtained one false negative (less than 50%).

Trial	**Classification**	**1 - False Negative**
1	**99.89 %**	
2	**99.90 %**	
3	**76.33 %**	
4	**84.07 %**	
5	**18.68 %**	✓
6	**99.99 %**	
7	**96.70 %**	
8	**98.79 %**	
9	**97.87 %**	
10	**98.97 %**	

VII. Conclusion

A probabilistic approach for 3D grasping movement recognition was proposed. A useful calibration between stereo camera and magnetic tracker device for human motion capture tasks was presented. After demonstrations of top and side grasp movements, features from these 3D trajectories are extracted, e.g., changes in direction (curvatures) and hand orientations. The learning phase is based on histogram techniques to quantify the probability distributions. A simple dynamic Bayesian network is used for classification. The preliminary results presented suitable classification of reach-to-grasp movements. However, a major evaluation has to be done with much more trials. A deep test and analysis will be carried out in the future to confirm the robustness of the recognition phase. This approach can also be used for gesture recognition. Other features such as velocity, backward and forward movements will be addressed in the future to improve the performance of the classification method.

References

[1] J. Rett and J. Dias, “Human-robot interface with anticipatory characteristics based on Laban Movement Analysis and Bayesian models”, *Proceedings of the IEEE 10th International Conference on Rehabilitation Robotics* (ICORR), 2007, pp. 257-268.

[2] E.-J Holden and R. Owens, “Representing the finger-only topology for hand shape recognition*”, In Machine Graphics & Vision International Journal*, Volume 12 , Issue 2, 2003, pp. 187 – 202

[3] R. Urtasun**,** D. J. Fleet and P. Fua. “Gaussian Process Dynamical Models for 3D people tracking”, *In Conference on Computer Vision and Pattern Recognition* **(***CVPR***)** *New York,* June 2006.

[4] S. Schaal, J. Peters, J. Nakanishi, and A. Ijspeert, "Learning Movement Primitives," in *International Symposium on Robotics Research (ISRR2003), Springer Tracts in Advanced Robotics*. Ciena, Italy: Springer, 2004.

[5] Odest Chadwicke Jenkins and Maja J Mataric´, "Performance-Derived Behavior Vocabularies: Data-driven Acqusition of Skills from Motion", *International Journal of Humanoid Robotics*, 2004.

[6] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. “Conditional models for contextual human motion recognition”, In *IEEE International Conference on Computer Vision*, volume 2, 2005, pp. 1808-1815.

[7] Motor Representations for Hand Gesture Recognition and Imitation. Manuel Lopes and José Santos-Victor. IROS Workshop on Robot Programming by Demonstration. Las Vegas, USA, 2003.

[8] D.R. Faria and J. Dias, "Hand Trajectory Segmentation and Classification Using Bayesian Techniques", *Workshop on Grasp and Task Learning by Imitation 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France - Sept, 22-26, 2008, pp. 44-49.

[9] Polhemus Liberty Electromagnetic Motion Tracking System: http://www.polhemus.com/?page=Motion_Liberty.

[10] Videre Design Stereo Camera STH-MDCS3-9cm: http://www.videredesign.com/vision/sth_mdcs3.htm.

[11] T. Svoboda, D. Martinec and T. Pajdla. “A convenient multi-camera self-calibration for virtual environments”. *PRESENCE: Teleoperators and Virtual Environments*, , August 2005. MIT Press, pp 407-422, 14(4).

[12] K. S. Arun , T. S. Huang , S. D. Blostein. “Least-squares fitting of two 3-D point sets”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.9 n.5, Sept. 1987, p.698-700.

[13] 5DT Data-Glove: http://www.5dt.com/products/pdataglovemri.html.

[14] D. R. Faria and J. Dias. “3D Hand Trajectory Segmentation by Curvatures and Hand Orientation for Classification through a Probabilistic Approach***”****. In Proceedings (to appear) of The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS’09, St. Louis, MO, USA.

[15] D. C. Knill and A. Pouget, “The Bayesian brain: the role of uncertainty in neural coding and computation,” *TRENDS in Neurosciences*, vol. 27, 2004, pp. 712–719.