Grasping Movements Recognition in 3D Space using a Bayesian Approach

Diego R. Faria, Hadi Aliakbarpour and Jorge Dias

Abstract — **In this work we present grasping movements recognition in 3D space. We also present the idea of a database of different sensors data for different scenarios of grasping and handling tasks for our future works. Multi-sensor information for grasp tasks require sensors calibration and synchronized data with timestamp that we start 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 the reach-to-grasp movement. Observing the reach-to-grasp movements of different subjects we perform a learning phase based on histogram using the segmentation data. Based on a learning phase is possible to recognize the grasping movements applying Bayes rule by continuous classification based on multiplicative updates of beliefs. We developed an automated system to estimate and recognize two possible types of grasping by the hand movements performed by humans that are tracked by a magnetic tracking device [9]. 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 or object displacement.**

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 activate with determined tasks. Investigation about human trajectories is useful to analyze the hand shape, pose and velocity, that is, 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 also 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 application is a 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. A proposed method is presented in this work for hand shape representation technique that characterises the finger-only topology of the hand using cepstral coefficients. Techniques of speech signal processing are used for that. The 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 which some of them uses different sensors data towards to find 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 a trajectory to find 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 axis 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 reached undesired results as false negative and recognition of the trajectory with low probability.

III. Sensors Database, Possible Scenarios and Applications

We intend to create 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 grasp movements tracked by 3D points (magnetic tracking system), images sequences (monocular and stereo), fingers flexure during the movement (data-glove), force applied in the object during the manipulation (tactile sensors) and points of interest trough a subject sight (eye tracker).

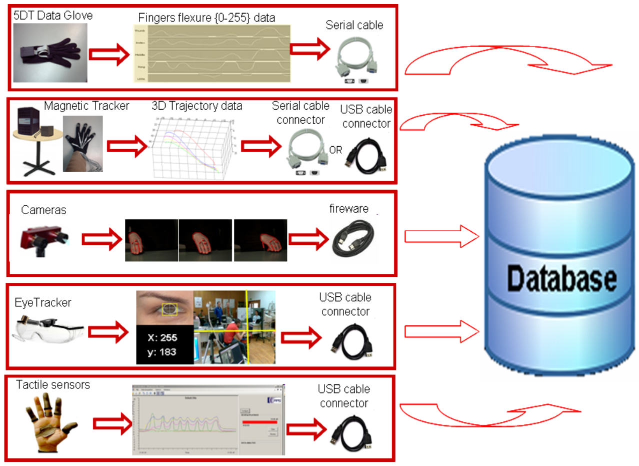

Fig.1. Sensors Database acquired in different grasping scenarios.

Depending on the scenario with a specific task more than one sensor will be used. For that, the sensors data that will be 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. For some grasping or manipulation applications are necessary the use of more than one sensors to complement each other. To work quite well with the data, a calibration between the sensors is needed.

A. Sensors Calibration Step

For some scenarios that are acquired human movements, sensors like magnetic tracker to acquire 3D positions and orientation are used. Images sequences for movements tracking are very useful and the most common used way to track human hands and face for human-robot interaction. 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 reproject the 3D points of the tracker device referential in the image plane (stereo camera referential) and vice-versa. The usability of this calibration is for a future work of 3D object shape representation integrating stereo camera and a glove with 5 sensors of tracker device in each fingertip for object manipulation acquiring its 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 referential, so that we can use the strategy of use a white tape on the sensor and then we can recognize this marker in the image getting the 3D point acquired after the camera calibration (Fig.2). We collected 30 images (left and right), they were acquired at same time of the 3D point from the tracker device sensor in different positions and orientations. The tracker sensor was attached at a tripod on a red piece of paper for easy displacement and easy localization in the image. This idea is originally inspired from auto calibration method between multi-cameras by Svoboda in [11] where was used a laser pointer to get different views point.

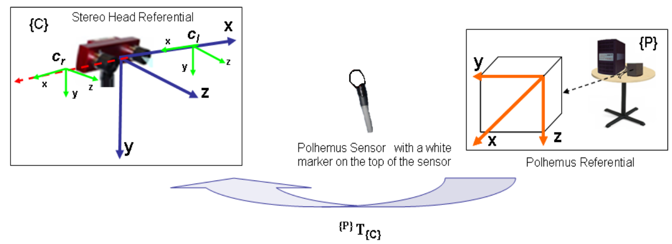

Fig.2 Calibration strategy: Using a white tape on the sensor we can get the 3D point of the sensor in the tracker device referential frame and in the camera referential frame.

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, ${}^{c}p=\{{}^{c}p_i \mid i=1,\ldots,N\}$ and ${}^{p}p=\{{}^{p}p_i \mid i=1,\ldots,N\}$ we compute the following equation to acquire a 3D point from a {P} to {C}:

$$ {}^{c}p={}^{p}R_c\,{}^{p}p+{}^{p}t_c \qquad (1)$$

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 $\boldsymbol{R}$ and $\mathbf{t}$ 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 at a tripod on a red piece of paper and the reprojection of its 3D point is represented as a yellow point in the image.

Table 1 shows the average reprojection 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.

	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

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

Fig.3. Reprojection of the 3D point (yellow color point inside of the circle) of the magnetic tracker in the image plan. The magnetic sensor with a white tape is attached at a tripod on a red piece of paper.

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

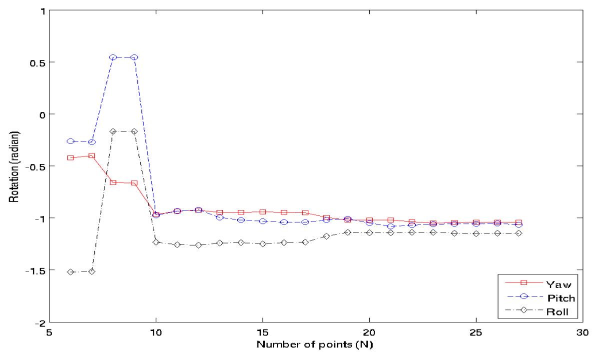

Fig.4. Evolution of the rotation matrix estimate by the calibration method according to the number of points used in the approach.

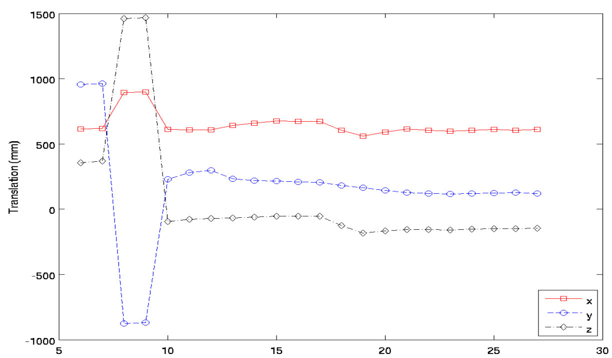

Fig.5. Evolution of the translation matrix estimate by the calibration method according to the number of points used in the approach.

Others sensors require individual calibration like 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 $\boldsymbol{R}$ is the dynamic range; $\boldsymbol{V_{max}}$ is the maximum output value (flexed hand) and $\boldsymbol{V_{min}}$ is the minimum output value (flat hand).
A normalization process is necessary and for that $\boldsymbol{R}$ is used, for example, lets work through the thumb the $\boldsymbol{V_{min}}$ and $\boldsymbol{V_{max}}$ for the thumb finger are 40 and 206 respectively , so that $\boldsymbol{R}$ = 166 . To scale the measured values across the full $\boldsymbol{R}$ 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; $\boldsymbol{M}$ is the measured value, for the thumb, $\boldsymbol{M}$ = 40 after normalization becomes in $\boldsymbol{N}$ = 0, when $\boldsymbol{M}$ = 206, after normalization becomes $\boldsymbol{N}$ = 255.

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.

B. *Possible Scenarios and Applications*

Through multi-sensors information, we can combine cues for a better understanding of how the human achieves grasping. Some questions can be posed related to how humans perform the grasp (Fig.6).

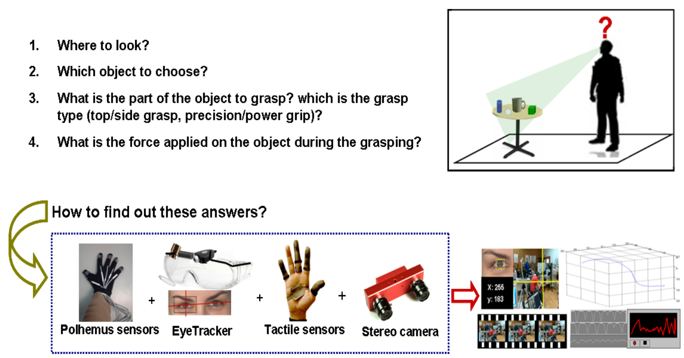

Fig.6. Some questions of how to achieve the grasping and a possible answer to solve this problem trough multi-senor cues.

For each scenario, a specific task concerning the research necessity can be determined. Several goals can be reached, some of them are trajectory analysis (reach-to-grasp) and hand pose estimation before the manipulation, analysis of force applied in the object during the manipulation, estimation of possible contact points on the object given its shape (cylinder, box, sphere, etc.). Other tasks like object manipulation and object displacement will be useful to endow a robot with skills to perform this goal.

Complex tasks can also be reached like estimation of the object that will be chosen to be grasped and the estimation of possible contact points on the object analyzing the person sight (trough eye-tracker device) and also by integration of multisensory information. Using Eye Tracker device we can track where the person is looking at and with this information we can try to know the contact points, that is, which part of the object the person will touch analyzing the direction of person's eyes. The hand trajectory until the object (reach-to-grasp) and the hand pose orientation can be verified trough the tracker device sensors. The force applied in the object during the grasping task can be analyzed trough the tactile sensors. Combining information of tracker device sensors on the fingertips and the fingers flexure given by a data-glove we can analyze the hand shape during the grasping. Object information can be acquired by stereo camera and markers on the object. These tasks are our work plan for future work.

IV. Scenario and Context

For this application we are using the same scenario proposed in previous work [8]. We have used the Polhemus Liberty tracking device [9] to track the trajectories performed by humans. We have attached five sensors in a glove to acquire the 3D hand trajectory allowing us analyzing the fingers behaviors during their journey to the object. Another sensor was placed on the 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 used in our application for classification: Top-Grasp and Side-Grasp (Fig.7). Fig.8 shows our current scenario and configuration for this application. We intend to estimate and recognize the grasp type in 3D space trough the movement performed by a human being.

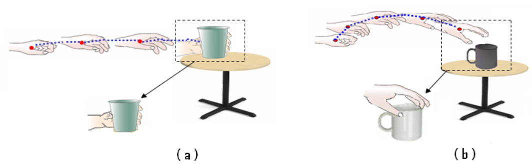

Fig.7. (a) – Side-grasp; (b) Top-grasp.

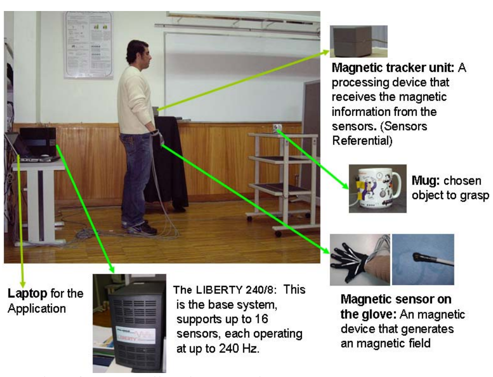

Fig.8. Scenario of our application: environment setup.

V. Trajectories Segmentation

The segmentation process is based on curvatures detection in 3D space. We have used the cylindrical coordinate system (***r, θ, h***) for that. The detected features are the possible directions combining ***h*** and ***θ*** information like *up, down, left, right, up-left, up-right, down-left, down-right and non-movement*. The equations for these features detection are described in [14]. Another segmentation step was used, the hand orientation along the trajectory by approximating the hand plan using 3 sensors on the fingertips (index, middle and ring finger). It is created 2 probabilities tables for each trajectory, one for the curvatures detection and another for hand orientation. The segmentation process is performed in each part of a normalized trajectory (all trajectories are normalized to the same scale). We have split the trajectory in 8 parts to find out the features in each one. For more information about these proposed methods of segmentation see [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]. Our learning phase is based on histogram techniques.

A. *Grasp Learned Tables*

In the learning phase is analyzed all probabilities tables of our dataset (hand orientation and curvatures), i.e. the results of the segmentation step. 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 | G D)***. The same rule is used for hand orientation learning, so that we have ***P(O| G D)*** where ***O*** represent all possible hand orientation (top or side).
The learned table is a mean histogram calculated from all top grasp and all side grasp probability tables. For more details see the work presented in [14]. Fig.9 shows our learned grasp tables for curvatures detection and Fig.10 shows the learned tables for hand orientation detection.

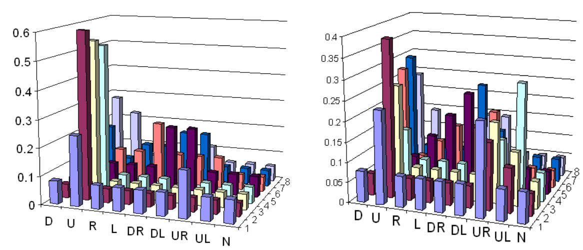

Fig.9. Left image represents the top-grasp curvatures learned table and right image side-grasp learned table.

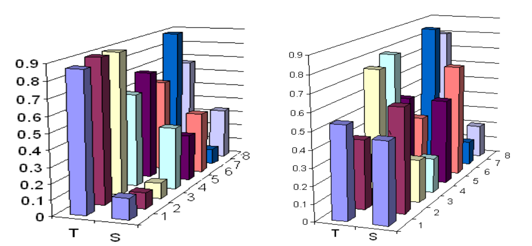

Fig.10. Left image represents the top-grasp hand orientation learned table and right image side-grasp learned table.

B. *Classification Model Using Bayesian Techniques*

Bayesian classification models have already proven their usability in gesture recognition systems as we can see in the work presented in [3], [8] and [14]. Based on this study we present 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 determined hand displacement is updated the probability of each type of grasp, i.e. the application informs us which grasping is more probable to happen by the highest probability between top and side grasp variables. Assuming the trajectory that is being performed has size 1 due we knowing the trajectory size a priori, i.e. we have the initial hand position and the mug position given by the sensors, then at each hand displacement is showed the probability of each grasp type. In our work [14] we have presented two Bayesian equations for trajectory classification, one using curvatures and another one using hand orientation. It was compared the results of each classification, by curvature and by hand orientation showing which of them achieved better results. In this work we present another Bayesian model integrating the two kinds of features in this model, that is, the two likelihood is used in the same model.

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) of the learned table.

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 their priors $P(G)$ 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 variables (priors) G are a uniform distribution and during the classification their values is updated applying Bayes rule shown in the next equations that is applied for top and side grasp:

$$P(g_{top} \mid c_{k+1}, o_{k+1}, i) \\ \alpha\ P(c_{k+1}, i \mid g_{top})\, P(o_{k+1}, i \mid g_{top})\, P(g_{top}) \tag{4}$$

$$P(g_{side} \mid c_{k+1}, o_{k+1}, i) \\ \alpha\ P(c_{k+1}, i \mid g_{side})\, P(o_{k+1}, i \mid g_{side})\, P(g_{side}) \tag{5}$$

$$P(g_{top} \mid c_{k+1}, o_{k+1}, i) \\ = \frac{P(c_{k+1}, i \mid g_{top})\, P(o_{k+1}, i \mid g_{top})\, P(g_{top})}{\sum_j P(g_j \mid c_{k+1}, o_{k+1}, i)} \tag{6}$$

$$P(g_{side} \mid c_{k+1}, o_{k+1}, i) \\ = \frac{P(c_{k+1}, i \mid g_{side})\, P(o_{k+1}, i \mid g_{side})\, P(g_{side})}{\sum_j P(g_j \mid c_{k+1}, o_{k+1}, i)} \tag{7}$$

First we compute (equations 4 and 5) the probability of all possible G (top and side grasp). The next step is the normalization, each variable $\mathbf{g}_{top}$ and $\mathbf{g}_{side}$ is normalized (equations 6 and 7). In the equations 6 and 7 the variable j is an index that represents all possible grasp (top or side). The main idea of the online classification is when someone is performing a trajectory, in the first hand displacement is applied the segmentation process to acquire the curvatures and hand orientations, after that is chosen the curvature and hand orientation with higher probability. The curvature with higher probability is multiplied by the correspondent curvature in the learned histogram (by same type of curvature in the same part of the trajectory) and this represents one of the likelihoods in the Bayes rule. The same is done with the hand orientation; the feature with higher probability is multiplied by the same correspondent feature in the learned histogram. The likelihood using curvatures and the likelihood using hand orientation are multiplied and then the result is multiplied by the prior (in the first iteration it is a uniform distribution). These steps are done until the last hand displacement (end of the trajectory).

The posterior probability of a current hand displacement becomes the prior for the next displacement. We formulate the equation as recursive way. Assuming that each hand displacement we can find new curvatures and new hand orientation then we can express the online behaviour by using the index k 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 reaching a certain threshold (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. *Experimental Results of Estimation and Recognition*

Fig.11 shows a side-grasp trajectory performed by a subject and table 2 shows the answer of our application along this trajectory classifying and recognizing it by the Bayes rule using the curvatures and hand orientation information. It shows the probability of being top or side grasp for each part of the trajectory. Using this Bayesian model integrating trajectory curvatures and hand orientation information in the same model we have reached a good recognition of the trajectory shown in Fig.11.

Table 3 shows 10 trials performed by different subjects for side grasp trajectories for result analysis. After 10 trials we can say that the proposed method integrating the curvatures and hand orientation features in the same model for the recognition step reached good results. A false negative value (trial 5) happened due the side-grasp trajectory be similar to the top-grasp concerning its learned features. A deep study and tests performing much more trials need to be done with this classification model for better analysis testing both top and side grasp trajectories.

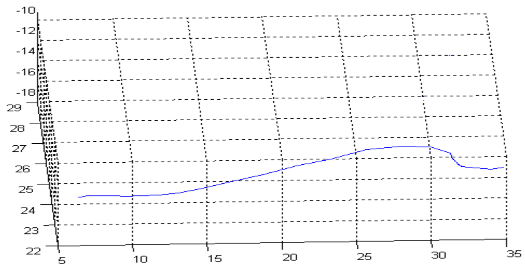

Fig.11. Side-grasp trajectory (raw data measured in inches).

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.2. Result of estimation and recognition of the trajectory shown in Fig.11. It was recognized as side grasp with 99.99%.

Trial	**Recognition Probability**	**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 %**	

Tab.3. Result of 10 trials of Side-grasp. In 10 trials just one false Negative (less than 50%) was reached.

VII. Conclusion

We have developed an application for recognition of grasp movements. Some ideas for multi-sensors information to analyze how the humans achieve grasping were presented for implementation in a future work. A useful calibration between stereo camera and magnetic tracker device for some human motion capture tasks were presented. A dataset of reach-to-grasp movements were created to be used in a learning phase based on histogram. A Bayesian model for classification and recognition of the grasp movement is presented. It uses two independent types of features detected in a segmentation process on the reach-to-grasp movements. The results show a good performance of the recognition step using the Bayesian approach presented. A deep test and analysis with much more trials are necessary to confirm the robustness of the recognition phase. This recognition method can be used also for gesture recognition tasks and for other kind of trajectories classification. Other segmentation information like velocity, backward and forward movements can be also used to test the classification method. All these information acquired by segmentation process can be used as initial step before the manipulation in robotics field. These actions can be learned and mapped to a robot perform these human movements in different grasp tasks.

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.