

# Probabilistic Classification of Grasping Behaviours using Visuo-Haptic Perception

S. Jafar Hosseini<sup>1</sup>, Diego R. Faria<sup>1</sup>, Jorge Lobo<sup>1</sup> and Jorge Dias<sup>1</sup>

<sup>1</sup>ISR - Institute of Systems and Robotics,  
DEEC, FCT, University of Coimbra, Portugal  
{jafar, diego, jlobo, jorge}@isr.uc.pt

**Abstract.** This paper presents a novel approach to visuo-haptic perception of grasping/manipulative tasks. The proposed approach is founded on a hierarchical Bayesian model which integrates the visual information with the haptic data to reach a reasonable percept of what is happening in grasping tasks. The primary goal of the approach is to identify what type of grasping behaviour is being performed by the human subject, and as a secondary goal, to simultaneously assess the quality of the respective grasping behaviour. For a simple set of grasping behaviours defined in this paper, preliminary experimental results indicate that the proposed approach could result in a robust and efficient perception of grasp behaviours.

**Keywords:** Visuo-Haptic Perception, Grasping Behaviours, Hierarchical Bayesian Models (HBMs).

## 1 Introduction

A key factor to achieve robust perception is the combination and integration of multiple sources of sensory information. Perceptual mechanisms, either technical or biological, rely on multiple sources to overcome shortcomings and noise of single sensory modalities, so that they can properly ‘perceive and act’. Many works from the cognitive science community have been focused on how humans achieve robust perception through the integration of information from multiple sensory modalities. In the case of grasping/manipulative tasks, humans make use of different sensory modalities, namely the visual and tactile senses as well as proprioceptive. These affect the ways that we accomplish manipulative movements on various objects. In fact, in order to grasp and manipulate an object with our hands, we need information about its size, shape, texture and mass. This information is in many cases conveyed by the visual and haptic senses.

In this work, we developed a visuo-haptic perception approach which combines visual data from a vision system observing the scene and haptic data from a tactile sensor on the hand, both acquired during grasping demonstrations performed by human subjects. The proposed approach is aimed at modelling grasp behaviours that occur during handling/manipulative tasks, by means of a probabilistic framework. This paper is organized as follows: section 2 will discuss the state of the art behind this paper. Section 3 describes how this work can contribute to the main theme of

conference. In section 4, the proposed approach and methods involved will be presented in detail. Next, section 5 presents the experimental setup used, experimental sessions and the preliminary results. In the end, conclusions and future work are presented.

## **2 Contribution to Technological Innovation for Value Creation**

This work falls within the scientific discipline “Autonomous robotics, specialization in robotic grasping and manipulation”. The major contribution of the work could be broken down in two points: 1- Investigating grasp behaviours from a new viewpoint by introducing an innovative visuo-haptic feature descriptor; 2- Proposing a hierarchical Bayesian-based framework to model grasp behaviours. The results derived from this work could be employed to enable Learning by Demonstration (LbD) frameworks to reproduce and replicate human like dexterity onto humanoid hand hardware. Enabling robots with more dexterity is a key factor for value creation, for instance in assisting the aging human population in household tasks.

## **3 Related Literature**

Many objects in our world can be picked up and freely manipulated by relying on information about an object from both the visual and haptic systems [1]. Most often, haptic perception refers to the extraction of information about an object’s properties through exploratory actions usually of the hands and arms [2]. While the importance of vision for object recognition and action planning has long been recognized [3], the role of the haptic sense for object recognition and manipulation has received considerably less attention. Although visuo-haptic perception has been extensively investigated by neuroscience researchers, it lacks adequate extension to applications in robotics research, more specifically, autonomous robotic grasping/manipulation. There is, however, some previous work which is aimed at merging visual and tactile data into a robust framework to provide the grasping robot with the required extent of autonomy. Most previous work combining vision and touch has been concerned with perceptual issues. One focus has been the use of touch sensing to supplement visual information for object recognition or exploration [4]. Rucci and Bajcsy [5] have examined coordination of active visual and tactile perception, and another effort developed representations for the integration of visual and tactile spatial information [6]. It’s worth noting that a number of the previous works employ vision system and force/tactile sensor mainly to the problem of design and development of feedback control for the humanoid hand. Jae S. Son proposes a generalizable framework of behavioural primitives for tactile and visual feedback control [7]. He focuses on the complementary roles of vision and touch. On the other hand, some other works consider the problem of segmentation and classification of demonstrator’s hand motion primitives using visual and haptic information, which is then applied to the purpose of Learning by Demonstration (LbD) frameworks [8].

## 4 Problem Formulation and the Proposed Approach

### 4.1 General Concepts of the Approach

The most significant aspect that makes a distinction between this work and other previous works is that we adopt a new insight into “grasping behaviours”. While performing grasping/manipulative tasks, the strategies that humans adopt to grasp an object could be expressed in terms of behaviour. In this work we consider a new specific set of “grasping behaviours” to which specific definitions will be assigned. Various kinds of behaviour may be observed during reaching/grasping movements carried out by different subjects. First, we need some meaningful criteria to characterize a grasp behaviour. The idea is to categorize different grasping actions from various aspects such as the type of grasp, the speed of movement, the force levels applied on the object’s surface, the flexure levels of fingers, the object being grasped, and so on. Haptic data carries useful information that can be used to supply low level features needed to Fig. out the force levels and the flexure levels of the fingers. In addition, visual data can be employed to help to capture the motion, and estimate the position of the hand and/or object. In turn, the tracking results provide important features by which the velocity of the movement can be computed. For simplicity, we further denote the force levels, flexure levels, movement velocity and grasping behaviour by force, flexure, velocity and behaviour, respectively. So, we will take advantage of these three kinds of features obtained from haptic and visual data to model the grasp behaviours shown by the subject during manipulating/grasping of an object. In fact, a hand motion descriptor can be constructed over these three features so as to define the various behaviours. However, in this work, we will only attempt to classify some types of behaviour. We will also consider the effectiveness of the behaviours, i.e., if they are successful in performing the desired task. A reaching/grasping process which is fine and stable is referred to as a normal behaviour chosen as reference for evaluation of other behaviours. The type of behaviour exhibited by humans depends on several conditions induced by the external environment which humans are interacting with. So, humans respond to each specific stimulus with a particular behaviour. Some of these conditions could be listed as: the properties of the object of interest like shape, size, weight and so forth; the position of the object with respect to the subject’s pose; lighting condition; subject’s intension of grasping; and so on. Considering the conditions enumerated above, we will only consider the following simple behaviours which often occur in everyday life:

1. Strong/sudden behaviour: Where the subject grasps an object hurriedly with much energy;
2. Light/sudden behaviour: Where the subject grasps an object hurriedly with a little energy;
3. Strong/sustained behaviour: Where the subject grasps an object slowly with much energy;
4. Light/sustained behaviour: Where the subject grasps an object slowly with a little energy.

The objective is to recognize these behaviours and at the same time, estimate their success rate. The reason we considered these particular behaviours is that they are not only commonly used, but also rich enough to perform an evaluation of the proposed approach. As discussed earlier, a probabilistic framework representing a hierarchical Bayesian model is built to combine the features force, flexure and velocity in order to cope with the visuo-haptic perception of how humans grasp an object.

## 4.2 Extraction of the Defined Features

The following paragraphs describe the key points on how to extract and estimate the desired features from the visual and tactile sensory data. In other words, the methods applied to label hand motions with respect to each feature are presented.

**Force:** We utilize and follow the strategy proposed by R. Martins [8]. He suggests that using tactile sensor information, a symbolic level generalization of manipulation tasks can be achieved. The tactile sensing device consists of 360 sensing elements (*Tekscan Grip System*) which are distributed along the hand palm and fingers surface. The sensing elements are grouped into 15 regions corresponding to different areas of the hand. A variable  $\mathbf{T}_i$  is assigned to each of these regions:  $\mathbf{T} = \{\mathbf{T}_1; \mathbf{T}_2; \dots; \mathbf{T}_{15}\}$ . The domain for each of those variables can be defined as  $\mathbf{T}_i \in \{\text{Low-Active}; \text{Medium-Active}; \text{High-Active}\}$ . The Low-Active, Medium-Active, High-Active define the level of activation of that region during the grasping task.

**Flexure:** During grasping/manipulative movements, the amount of flexion of each finger of the hand may vary from having no flexion to having a full flexion. It should be pointed out that a wide variety of hand configurations could arise from combination of different flexion levels of the fingers, which implies that each grasp type corresponds to a separate set of flexion levels as some of the hand configurations represent different grasp types. Therefore, grasp type as a subset of flexure could be regarded as a high level feature to describe behaviours. Each finger is composed of three joints, with the exception of the thumb that has two joints. Given the flexion amount of all joints from an instrumented glove (CyberGlove), an average value of flexion is then computed for each finger from its joints' flexion. As in force, the entire extent of flexion is divided into three levels named "Low-Flexed", "Medium-Flexed" and "High-Flexed", proportionally distributed over the whole data range, as done for the force.

**Velocity:** Let's assume that the 3D position of the hand is provided with the aid of some kind of motion tracker device (in our case, vision system). Once the 3D pose has been obtained, we are able to estimate the velocity of the movement by differentiating the distance travelled by the hand with respect to time. Actually, this feature gives the possibility of dynamic analysis for hand motions and supplements the static analysis obtained from flexure and force, although these could also be dynamically analysed, but it would lead to a much more complex model. Then, depending on the velocity, the hand motion is classified as either fast or slow. To do so, we check if the velocity is above or below a pre-defined threshold.

Now, we have the descriptor ready to encode the handling/grasping motions in terms of force, flexure and velocity.

### 4.3 Probabilistic-Based Approach

Any model of a real phenomenon is incomplete as there are always some hidden variables which cause uncertainty in the model. The more important the effects of the hidden variables, the more "noisy" the data is. Recently, cognitive modellers have begun to address hierarchical Bayesian models (HBMs) to tackle incompleteness. Hence, after all features are completely extracted, we create a probabilistic framework inspired by HBMs. For now, to simplify this framework and avoid much complexity, we only consider the index finger, which is the one most involved in handling tasks. This framework which is designed to model the grasp behaviours deals with parameters denoting force, flexure, velocity, object, success indicator and behaviour as shown in Fig. 1. The following equation represents the joint distribution for the proposed Bayesian model:

$$P(F1 \wedge F2 \wedge F3 \wedge X \wedge V \wedge S \wedge O \wedge T \wedge P \mid B) = P(F1 \mid P) * P(F2 \mid P) * P(F3 \mid P) * P(X \mid O \wedge T) \quad (1) \\ * P(V \mid S \wedge O \wedge T) * P(S \mid B) * P(O \mid B) * P(T \mid B \wedge O) * P(P \mid B \wedge S \wedge O \wedge T)$$

where  $F1$ ,  $F2$  and  $F3$  denote force of regions 1, 2 and 3 of index, respectively.  $X$ ,  $V$ ,  $S$ ,  $O$ ,  $T$ ,  $P$  and  $B$  denote flexure of index, velocity of index, success indicator, object being grasped, task context, the level of power (light or strong) and behaviour, respectively. Success indicator is used to distinguish between a normal behaviour that accomplishes the task and the other ones. As the grasping study is task-dependent, the variable  $T$  is added to indicate the task context in which a kind of grasping/manipulative tasks takes place. To find some of these probability distributions, learning phase experiments were performed over the learning data. Then, given the test data, the proposed probabilistic framework integrates force, flexure and velocity into a robust percept to classify grasp behaviours and judge how successful a generic behaviour is. To this end, Bayesian inference is executed by making pertinent questions to the model.

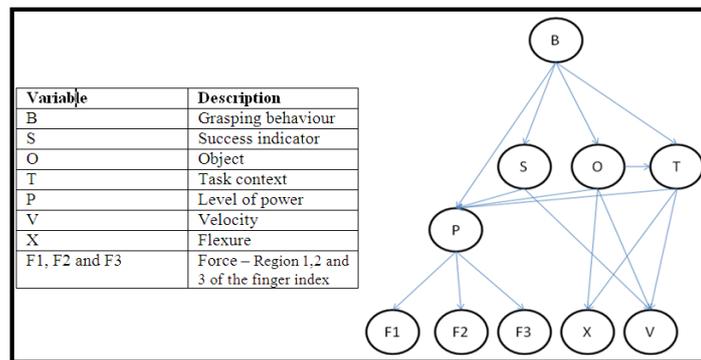


Fig 1. A hierarchical Bayesian network to model grasping behaviours

As can be seen from Fig. 1, this model is composed of three levels of abstraction. The low-level consists of features from visuo-haptic data. Success indicator, object and task context are placed at the mid-level, and behaviour is at the high-level. Notice that some of the variables have been assumed to be independent of one another. Nevertheless, it is not possible to guarantee that force and velocity don't influence

each other at all. Thus, it is indisputable to take into account the dependency hypotheses between force and velocity when needed. For example, some relation between force and velocity could be stated as: grasping a very heavy object involves exerting more force as compared to a normal object and also the high weight of the object usually causes the subject not to be able to manipulate it fast. As a result, there might be more dependencies among these variables, which will lead to a complicated hierarchical Bayesian network. Furthermore, this network would get still more complicated when taking account of all fingers and the interactions among them. In this paper we adopt a simpler model to validate the approach.

## 5 Experimental Setup and Results

To evaluate the performance and efficiency of the proposed approach, we tested it on datasets acquired by conducting some experimental sessions. To do the experiments, we benefited from the experimental setup constructed at our lab for the HANDLE project [9]. This setup consists of several sensory devices as illustrated in Fig. 2: Tekscan tactile device which generates force information; Polhemus motion tracker which provides the 3D position and orientation of the fingertips; CyberGlove which gives flexure information; Monocular camera; Stereo camera; and Kinect sensor.

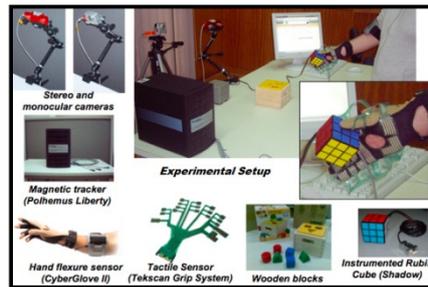


Fig 2. HANDLE experimental setup (Adapted from [9])

Among the sensors mentioned above, we used Tekscan tactile device and CyberGlove as haptic sensors and a stereo camera as visual sensor. Fig. 3 presents a sample example of learning histograms. The proposed approach was tested with two different objects as well as two different task contexts. Two subjects were asked to participate in both pickup-place and pickup-pour tasks with a Rubik cube or mug as shown in Fig. 4. Each dataset was sampled with a rate of 30 Hz and the three feature spaces were re-sampled at equal timestamps. The results of implementation of the proposed approach on two different test datasets, each of which represented a distinct behaviour, are shown in Fig's 5.a and 5.b. At each sample of dataset, the inference questions take as input  $F1$ ,  $F2$ ,  $F3$ ,  $X$ ,  $V$ ,  $O$  and  $T$  and subsequently output the probability distributions of  $B$  and  $S$ . Along the grasp trajectory, the results are updated since  $P(B|F1^{\wedge}F2^{\wedge}F3^{\wedge}X^{\wedge}V^{\wedge}O^{\wedge}T)$  at the current sample is used as prior distribution for the next sample and consequently the final results are those obtained in the last sample. These preliminary results show that the proposed approach successfully classifies the observed grasping behaviours.

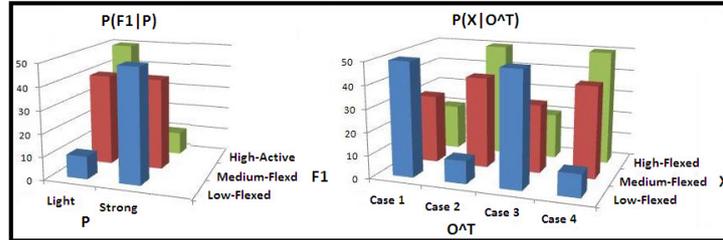


Fig 3. A sample example of learning histograms



Fig 4. Subjects performing grasping tasks

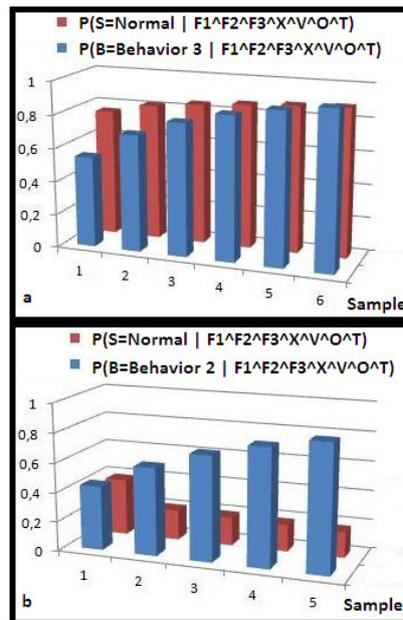


Fig 5. a) A subject performed pickup-pour with the mug by applying behaviour 3 (Strong/sustained) which is fairly normal and successful (The time interval between every two samples is 990 ms). b) A subject performed pickup-place with the Rubic cube by applying behaviour 2 (Light/sudden) which is rather abnormal and not successful (The time interval between every two samples is 330 ms)

## 6 Conclusion and Future Work

In this work, we have developed an approach to modelling the grasping behaviours through visuo-haptic perception. According to this approach, a motion descriptor is built over haptic and visual data and a probabilistic framework is then adapted to it. Across this framework, haptic data is complemented with visual data to get a strong perception in order to classify grasp behaviours and assess their effectiveness. Observing the preliminary results, we can see that the proposed approach proved consistent in the classification, in spite of the noise and perturbation inherent in the model.

This work is a starting point, motivating us to extend the idea to more comprehensive approaches, which we are going to address in our future work. Several goals will be pursued: improvement of the current model in a way that integrates a novel descriptive language called Laban-based Hand Motion Analysis (LHMA) into a more complex HBM, which will take into consideration all the fingers, instead of just the index finger, and characterize more behaviours; mapping of the human-like grasp behaviours to real humanoid hands.

As mentioned above, the results derived from this work could contribute to human-like dexterity onto humanoid hand hardware. For instance in assisting the aging human population in household tasks, having more dexterous robots is a key factor.

**Acknowledgments.** The research leading to these results has been partially supported by the HANDLE project, which has received funding from the European Community's 7th Framework Programme under grant agreement ICT 231640; and by the Institute of Systems and Robotics at Coimbra University.

## References

1. Newell, F.N.; Bühlhoff, H.H. & Ernst, M.O. *Cross-modal perception of actively explored objects. Eurohaptics 2003 Conference Proceedings*, 2003, 291-299.
2. Henriques DY, Soechting JF (2005) Approaches to the study of haptic sensing. *J Neurophysiol* 93:3036-3043
3. von Helmholtz H (1867) *Handbuch der physiologischen optik*, vol. Bd. 9. Voss, Leipzig
4. M. Boshra and H. Zhang, A constraint-satisfaction approach for 3D visionitouch-based object recognition. *IEEWRSJ International Conference on Intelligent Robots and Systems*. vol. 3, pp. 368-73. 1995.
5. M. Rucci and R. Bajcsy. Learning visuo-tactile coordination in robotic systems. *IEEE International Conference on Robotics and Automation*. vol. 3, pp. 2678-83. 1995.
6. N. I. Rafla and E L. Merat. Vision-taction integration for surface representation. *IEEE International Conference on Systems Engineering*. pp. 519-22. 1990.
7. Jae S. Son, R.D. Howe, J. Wang, and G.D. Hager. Preliminary results on grasping with vision and touch. In *In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS96*, volume 3, pages 1068–1075, Osaka, Japan, November 1996.
8. R. Martins, D. R. Faria, J. Dias - Symbolic Level Generalization of In-hand Manipulation Tasks from Human Demonstrations using Tactile Data Information. *Workshop on Grasping Planning and Task Learning by Imitation*, Taipei, Taiwan - October 2010.
9. HANDLE project: Developmental pathway towards autonomy and dexterity in robot in-hand manipulation, EC FP7 ICT-236410, [www.handleproject.eu](http://www.handleproject.eu), 2011.