

# Using experience for assessing grasp reliability<sup>1</sup>

Antonio Morales<sup>1</sup>, Eris Chinellato<sup>1</sup>, Andrew H. Fagg<sup>2</sup>, and Ángel P. del Pobil<sup>1</sup>

<sup>1</sup> Robotic Intelligence Laboratory, Universitat Jaume I, Castellón, Spain  
morales@icc.uji.es, eris@icc.uji.es, pobil@ieee.org

<sup>2</sup> Laboratory for Perceptual Robotics, University of Massachusetts  
fagg@cs.umass.edu

**Abstract.** Manipulation skills are a key issue for a humanoid robot. Here, we are interested in a vision-based grasping system able to deal with previously unknown objects in real time and in an intelligent manner. Starting from a number of feasible candidate grasps, we focus on the problem of predicting their reliability using the knowledge acquired in previous grasping experiences. A set of visual features which take into account physical properties that can affect the stability and reliability of a grasp are defined. A humanoid robot obtains its grasping experience by repeating a large number of grasping actions on different objects. An experimental protocol is established in order to classify grasps according to their reliability. Two prediction/classification strategies are defined which allow the robot to predict the outcome of a grasp only analyzing its visual features. The results indicate that these strategies are adequate to predict the reliability of a grasp and to generalize to different objects.

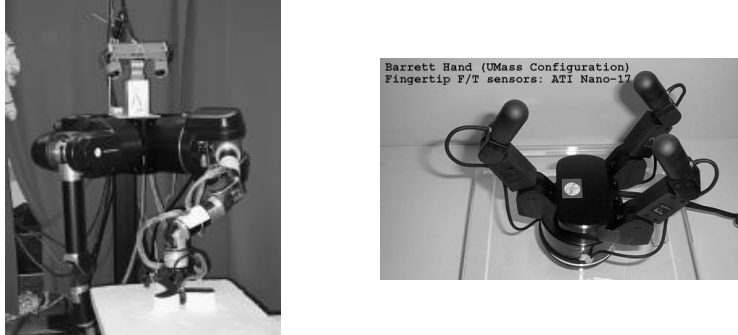
## 1 Introduction

The global behavior of a humanoid robot should emerge from the ability to execute simple tasks and combine them together in an intelligent way. The important skill we are interested in is that of being able to interact with objects, moving them with arm-like limbs. Specifically, we aim at improving the grasping skills of a robot interacting in a real, non-customized environment. We are interested in a visual-based grasping system that can provide the robot with the ability of grasping objects never seen before (and unmodeled) in real time. The robot should also be able to improve its performance by learning from its previous grasping experiences.

When a robot faces a certain practical problem, the task to be performed can be normally subdivided into lower-level tasks, until reaching basic actions. Conversely, these basic tasks can be put together to form more complex behaviors. Interesting to note, primates do not differ too much from this pattern of behavior. Indeed, prototypes of simple actions (both cognitive and practical ones) are coded in primate brains, and the way they connect originates higher

---

<sup>1</sup> Paper submitted to and accepted in Third IEEE International Conference on Humanoid Robots, to take place in Karlsruhe, Germany in October 2003.

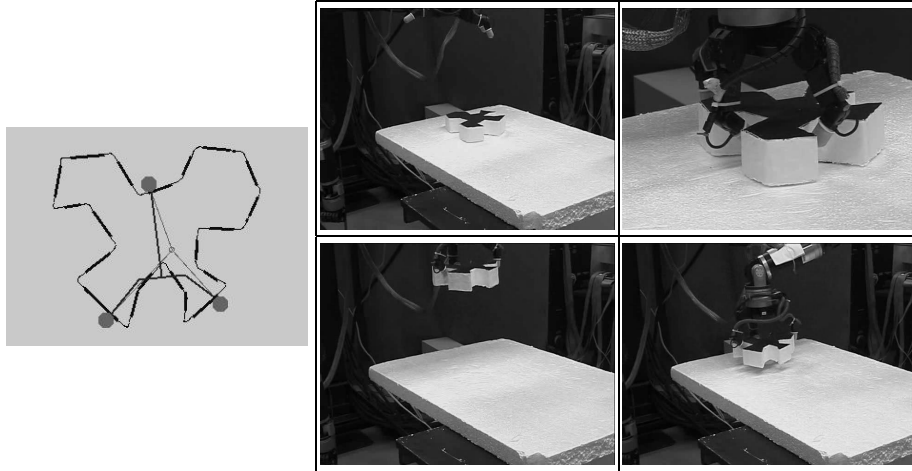


**Fig. 1.** The UMASS Torso. A humanoid robotic system developed in the Laboratory for Perceptual Robotics in the University of Massachusetts. On the right, a detail of one of the two Barret hands[7] that are part of the system.

level behaviors. In the case of grasping, specific areas of the posterior motor cortex are believed to contain a vocabulary of motor actions of different level (e.g. preshape the fingers for a precision grip) [18]. Such actions are selected and combined in different ways according to the task (e.g. push or grasp), the object shape and size, the timing of the action and other issues.

Another relevant aspect coming from neuroscience studies [9, 19] is that visual processes related to grasping actions in primates are different from visual processes having different goals. In fact, looking at an object with grasping purposes activates a peculiar neural pathway which is not active if grasping actions are not involved. This activation seems to represent a 'potential grasping action'. Some neurons in the anterior intraparietal area (AIP) are found to be active when grasping some particular objects, but also when looking at them with the purpose of grasping (and only in those cases). Some other neurons of the same zone are sensitive to the size or orientation of the objects. Therefore, the AIP area seem to encode the 3D features of objects in a way that is suitable to guide the movements for grasping them, which are stored in the premotor cortex. Not surprisingly, the AIP area does connect strictly to the areas of the motor cortex containing the motion primitives which have been mentioned above. Thus, as stated by [8], "object attributes are processed differently according to the task in which a subject is involved. To serve object-oriented action, these attributes are subjected to a 'pragmatic' mode of processing, the function of which is to extract parameters that are relevant to action, and to generate the corresponding motor commands".

In our approach to the grasping problem in a real environment, the emphasis has been on exploiting the use of on-line visual sensing, and the definition of visual attributes/features extracted from the object with the purpose of grasping. Interesting to note, and unlike the majority of the research, the grasping experience obtained with different objects is merged, meaning that experience related to a certain object is useful in order to predict the outcome of a grasp



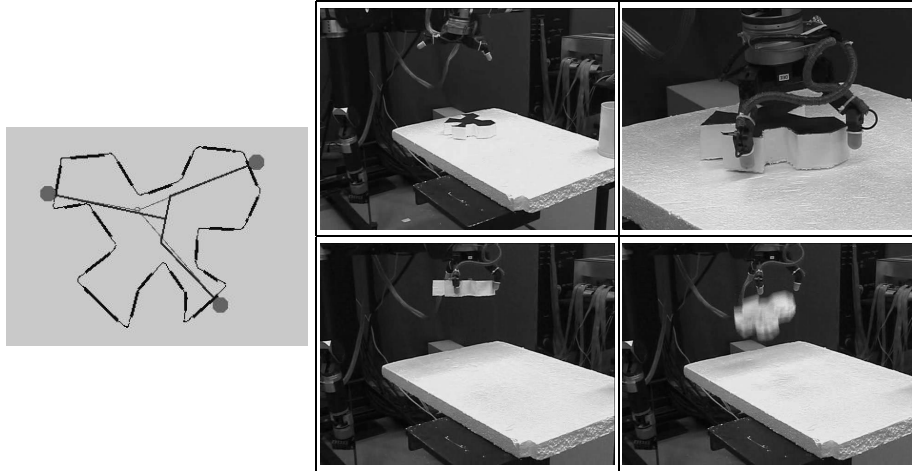
**Fig. 2.** Successful execution of a grip

performed on a different object. It contrasts to previous approaches which typically assume that a parameterized model of the entire object is known before grasp planning begins. When such a model exists, there is a well-founded corpus of analytical methods for grasp analysis and synthesis that are computationally expensive but could be applied off-line [1]. Even if the model is not previously available, such analytical methods could be applied after an exhaustive reconstruction phase from the visual input. However, this is not only too costly in terms of sensing and computation time to be feasible for real-world applications, but also in disagreement with the neurophysiological findings described above.

Instead of this, our system is able to compute a grip of a real unmodeled object using visual sensing in a grasp-oriented manner [13], and explicitly avoiding a full reconstruction followed by a classical grasp planning process. The repetition of grasping actions performed on different objects under different points of view provide the experience from which the robot learns the reliability of several possible grasps. In this paper we describe how the robot learns to predict the reliability of a grip starting from a number of grasping features and a set of practical experiments.

Interesting to note, and unlike the majority of the research, the grasping experience obtained with different objects is merged, meaning that experience related to a certain object is useful in order to predict the outcome of a grasp performed on a different object.

A key issue is to define the intrinsic features which characterize a grasping action starting from visual information of the object and proprioceptive information about the limb (and especially about the hand). Such features constitute the base for assessing a hypothetical grip in order to predict the possible outcome of its execution.



**Fig. 3.** Unsuccessful execution of a grip

We present in section 2 a set of nine features which characterize three-finger planar grasps. The features depend on both grasping-oriented visual information about the object and proprioceptive information about the hand geometry and kinematics.

Later we introduce a strategy devised within this feature space so as to predict the success of a particular grasp configuration. This strategy is based on a nonparametric classification technique, the *k-nearest neighbor rule* (Sec. 4.1). In section 5 we compare this technique with another one based on the use of a neural network scheme. These strategies are validated using real data gathered from experiments conducted on a humanoid system following the experimental protocol described in Sec. 3. Our results suggest that both methodologies are good enough to predict the reliability of a grasp within a reasonable error margin.

Finally, before going into more technical and detailed explanations, we show two illustrative examples of the problem we are facing. Figures 2 and 3 show the execution of two different grasping experiments. On the left side of the images is the hand configuration correspondent to the grasp to perform. The meaning of these figures are explained in more detail in section 2. On the right side of each figure are depicted four frames extracted from the grasping sequence, showing four successive instants of the action. From the first frames of each experiment is not obvious what the outcome of the grasping action will be. Though, the last frames show that, whilst the first grasping action (see fig. 2) was successful and the object put back down on the table, the same didn't occur in the second case (fig. 3), as the object was dropped a few seconds after the lifting movement. Thus, grips have different reliability, and our goal is to predict the most probable outcome of their execution.

## Experimental setup

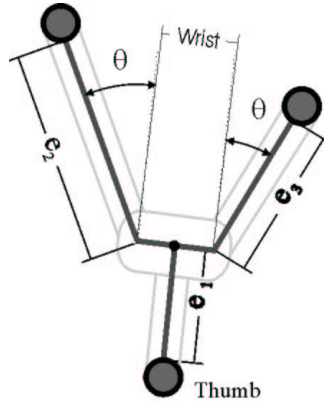
Our experiments have been implemented using the UMass Torso. This humanoid robot (Fig. 1) consists of two Whole Arm Manipulators (WAMs) from Barrett Technologies, two Barrett Hands, a BiSight stereo head, and a quadrasonic audio system. The WAMs are seven degree-of-freedom manipulators, whose kinematics are roughly anthropomorphic. The BiSight head consists of two cameras and ten controllable degrees of freedom (four mechanical and six optical).

An important component of the humanoid are the grippers, in our case two Barrett Hands (Fig. 1). Each of them is composed of three-fingers with a total of four controllable degrees of freedom. Each finger consists of two flexion degrees of freedom which are driven by a single motor. The fourth motor controls abduction/adduction of two of the fingers which are symmetrically placed on either side of the remaining finger, the *thumb*. When adducted, the two fingers flex in parallel to the thumb; when fully abducted, the two fingers flex in opposition to the thumb. This design is very compact, allowing for the motors and control CPU to be embedded into the wrist, but limits the possible placement of finger contacts. In the UMass configuration of the Barrett Hand, the finger tips are cylinders capped with hemispheres; each contains a six-axis force-torque sensor.

A high-level *control basis* provides a set of look/reach/grasp primitives upon which our reach and grasp sequences are implemented [4, 16]. The stereo vision system estimates the two-dimensional location of the target object on the table, and provides a monocular image for surface curvature analysis (see [14] for more details). Once a grip is selected (consisting of contact locations and a hand posture), the hand is preshaped and positioned above the object. When the grasp is physically established, the object is lifted and transported to a designated location.

## 2 Vision-Based Grasp Features

In the particular case of planar grasp determination, i.e. for objects resulting from the extrusion of a planar shape, we showed in [13, 14], how vision information can be used to select a set of feasible grips that meet certain stability criteria, including the particular kinematics of the three-fingered Barrett hand. This approach typically yields a large set of triplets of contact points, out of the infinite geometric possibilities. However, only one grip can be finally executed, and this choice can be mediated by further considerations such as the particular robot intentions, the task to be performed, additional reachability constraints, etc. An adequate characterization of the grips is called for in order to be able to predict their reliability and adjust practical aspects of the manipulation activity accordingly during the execution of the grasp and subsequent movements (e.g.: arm acceleration torques, etc). In this section we first describe some basic descriptors (see [13, 14] for more details) and then introduce the nine features used to characterize a grasp configuration. Fig. 4 shows a schematic representation of the kinematics of the Barrett hand relative to the palm on the palm.



**Fig. 4.** Barrett hand kinematics. The hand has a thumb and two opposing fingers that spread symmetrically along the axis defined by the thumb.

## 2.1 Grasp descriptors

The key properties of a grasp described using the following descriptors.

**Grasp regions.** The portions of the object contour where the three fingers are placed (indicated in bold in the images). These are modeled as short, straight segments and described by the coordinates of their extreme points.

**Contact points.** The three points where the fingers are supposed to touch the object, each lying on one of the three grasp regions ( $P_1, P_2, P_3$ ).

**Force directions.** The real force directions  $F_1, F_2, F_3$  exerted by the fingers of the Barrett hand are usually different from the ideal normal directions  $N_1, N_2, N_3$ .

**Force focus.** The intersection of the directions of the real forces  $C_C$ .

**Finger extensions.** The opening of the fingers ( $e_i$  in Fig. 4 and 6).

**Spread angle.** The spread angle ( $\theta$  in Fig. 1) of the opposing fingers.

In addition to these grasp descriptors, some parameters that summarize physical aspects of the environment are introduced for their use in the definition of the features in the following subsection.

**Friction coefficient ( $\mu$ ).** Estimate of the minimum possible friction between fingers and object;  $\mu = 0.4$  was used, which is a very safe setting for rubber-covered fingers.

**Positioning error threshold ( $\lambda$ ).** Safety value for possible finger positioning errors due to visual and mechanical imprecisions; it is set to  $\lambda = 2\text{cm}$ , larger than the maximum positioning error observed in the practical experiments (1.5cm).

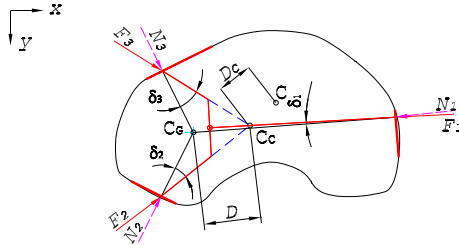
## 2.2 Feature definitions

The descriptors introduced in previous subsection somehow low level. We propose nine high-level features computed from the grip descriptors that try to measure different properties of each grip. Note that the inputs for their computation come only from the object contour extracted from the image along with the knowledge about the hand geometry. More details can be found in [2, 3].

All variables used in the different features are illustrated in Fig. 5, 6 and 7.

1. **FORCE LINE:** This feature [14] considers the deviations  $\delta_i$  of the real forces  $F_i$  from the ideal condition of being perpendicular to the contour at the grasp points. As the actual friction coefficient of the contacts between object and fingers is not known beforehand, the more the force deviates from the normal, the more the risk of the fingers sliding along the side of the object, due to a large tangential component of the applied force. Low deviations indicate low risk of instability:  $q_1 = \frac{4}{3}(\delta_1^2 + \delta_2^2 + \delta_3^2)/\arctan^2(\mu)$ .

2. **REAL FOCUS DEVIATION:** This feature measures the distance  $D$  between the focus of the ideal forces  $C_G$  and the real focus of the grip  $C_C$ . The feature is computed as  $q_2 = \frac{2D}{\eta\mu}$  where  $\eta$  is the maximum possible finger extension.



**Fig. 5.** Geometrical representation of the variables involved in the computation of features 1 ( $\delta_1$ ,  $\delta_2$ , and  $\delta_3$ ); feature 2 ( $D$ ,  $C_C$ , and  $C_G$ ); and feature 5 ( $D_C$ ).

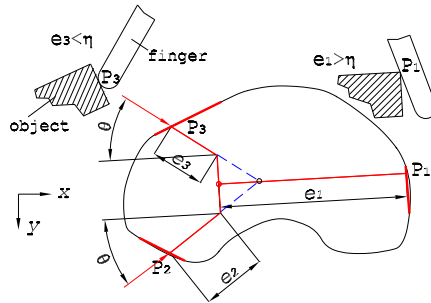
3. **FINGER EXTENSION:** If the fingers contact the object with different extensions, they touch the object in positions having slightly different distances from the surface, and they probably exert a torque out of the horizontal plane of the object. This feature estimates the risk given by differences in the finger extensions:  $q_3 = \frac{1}{\eta^2}((e_1 - e_2)^2 + (e_2 - e_3)^2 + (e_3 - e_1)^2)$ .

4. **FINGER SPREAD:** An equilibrated grip should have its three forces roughly equally separated by  $120^\circ$  angles [15]. Calling  $\theta$  the opening angle of the fingers

of the Barrett hand in opposition to the thumb,  $q_4 = (\frac{\pi}{6}/(\frac{\pi}{2} - \theta)) - 1$  for  $\theta > \frac{\pi}{3}$  or else 0. This implementation identifies situations that are at risk for violating force-closure, where the two opposing fingers have forces perpendicular to the thumb and facing each other.

5. REAL FOCUS CENTERING: This feature aims to measure the effect of gravitational and inertial forces endorsing grasps with short distances between the real focus  $C_C$  and the center of mass of the object  $C$ . The latter is the centroid of the two-dimensional shape described by the extracted object contour, assuming that the object has uniform mass distribution. The quality value is:  $q_5 = \frac{4D_C}{M_L + m_L}$ , where  $M_L$  and  $m_L$  are the sizes of the major and minor inertia axes computed for the shape.

6. FINGER LIMIT: When trying to grip large objects, there is a limit in the extension of the fingers. Due to the way the Barrett Hand grips objects, there is a finger extension value that, if overcome, causes the grip to shift from a fingertip grip to a fingerside grip to execute, which is more risky and less stable although still possible (see Fig. 6). Therefore, a threshold on the maximum optimal finger extension  $\epsilon$  has been set in order to avoid marginal contacts:  $q_6 = \epsilon_1 + \epsilon_2 + \epsilon_3$  where  $\epsilon_i = (\frac{e_i - \eta}{\lambda})^2$  if  $e_i > \eta$ , else 0.



**Fig. 6.** Geometrical representation of the variables involved in the computation of features 3 ( $e_1$ ,  $e_2$ , and  $e_3$ ); feature 4 ( $\theta$ ); and feature 6 ( $\eta$ ).

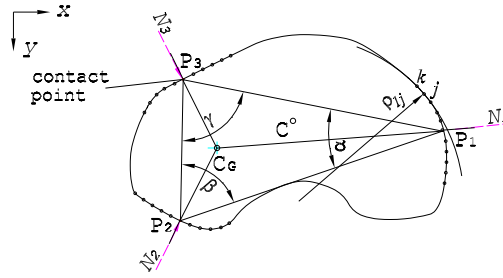
7. POINT ARRANGEMENT: Similarly to [10, 15], we assess the likeness of the grasp triangle to an equilateral one to obtain better grip balance. Each angle is compared with a  $60^\circ$  ( $\pi/3$  rad) angle typical of an equilateral triangle:  $q_7 = \frac{3}{2\pi}(|\alpha - \frac{\pi}{3}| + |\beta - \frac{\pi}{3}| + |\gamma - \frac{\pi}{3}|)$ .

8. TRIANGLE SIZE: The larger the area of the grasp triangle, the more stable a grip is [10]. This feature assesses the ability of the grip to resist the torques



generated by gravitational and inertial forces, whose magnitude is proportional to the weight of the object. The quality measure is  $q_8 = \frac{A}{4A_{S2}}$ , where  $A_{S2}$  is the area of the grasp triangle, and  $A$  is the area of the object.

9. CONTACT CURVATURE: A concave surface is a better place to put a finger for grasping purposes than a convex one [12]. This feature takes into account the curvature of the three grasp zones. All the points closer to the grasp point than the positioning error threshold are considered, and their local curvature values are summed. The sum is weighted by the actual distance of each point from the contact point, in a way that the more we approach the expected point of contact the more the local curvature value becomes influent on the total. The curvature  $\rho$  is positive for concavities, negative for convexities and 0 for planar zones. We define the overall grip quality as:  $q_9 = 3 * \psi - (\rho_1 + \rho_2 + \rho_3)$ , where  $\psi$  is the curvature threshold value, that is the best (most concave) possible curvature allowed for a contact point.  $\rho_i = \sum_{j=-k}^k (1 - \frac{|j|}{k}) * \rho_{ij}$ , with  $\rho_{ij}$  local curvature of a point that is at distance  $j$  along the contour from the point  $i$ . In practice the distance is measured in discrete steps. The maximum distance  $k$  depends on the positioning error  $\lambda$ .



**Fig. 7.** Geometrical representation of the variables involved in the computation of features 7 ( $\alpha$ ,  $\beta$  and  $\gamma$ ); 8; and 9 ( $\rho_{ij}$ ).

### 2.3 Feature normalization

Originally these features were designed as quality assessment criteria [2]. We claim that these criteria can be used to define a feature space with the purpose of characterizing robot grasps according to their reliability. Values close to 0 indicate more stable characteristics whereas larger values try to identify sources of instability. Moreover, all features are designed to have similar ranges. More precisely, they are implemented so as the best grips correspond with the lower

values, with an ideal theoretical best value of 0 (except for features 8 and 9, which are lower bounded to strictly positive values). Also, a preprocessing stage has been performed based on physical and numerical considerations. This consists of a normalization factor that is dependent on the distributions and ranges of each feature, so that a middle quality grip for a certain feature is expected to have a quality value of 1. When possible, the normalization value has been set according to physical aspects related to the feature. Otherwise, a value that is halfway between the best and the worst possible evaluations is the normalization value. Justifications for the normalization can be found in [2].

### 3 Methodology: Experimental protocol

A set of real objects has been built for this experiment. They are planar objects with a constant height made of an homogeneous material. Moreover, the colors of the objects have been selected to simplify the image processing. An important feature is that their shape is *unknown* for the system. The only programmed assumptions about the objects is that they are planar. The rest of the information, in particular the shape and location, is obtained from the images.

Moreover, in order to study the grasping performances in different circumstances several characteristics of the environment are tested. These are the weight of the objects and the friction coefficient. Two qualitative categories for each of both conditions are distinguished: heavy and light objects, and high and low friction. The different weight is obtained by making two different sets similar in appearance, but made of different material. Different contact friction is achieved by using a latex fingertip to envelope the fingers.

In order to perform the experiments, a single object is placed on a table within the robot workspace. Using the stereo-visual information the robot locates the object and computes a set of feasible grasp configurations. One of the configurations is selected, either manually by a human operator, or automatically by the robot, and executed.

If the robot has been able to lift the object safely, a set of stability tests are applied in sequence. These are aimed at measuring the stability of the current grasp. They consist of three consecutive shaking movements of the hand which are executed with an increasing acceleration. After each movement the tactile sensors are used to check whether the object has been drop off.

This protocol provides us with a qualitative measure of the success of a grasp. Thus, an experiment may result in five different reliability classes:  $E$  indicates that the system was not able of lifting the object at all;  $D$ ,  $C$ ,  $B$  indicate that the object was dropped, respectively, during the first, second, or third series of shaking movements; finally  $A$  means the object did not fall and was returned successfully to its initial position on the table. Hence, we define  $\Omega = \{A, B, C, D, E\}$  as the set of reliability classes.

The number of feasible grips that are computed for a single object is usually large, varying from several dozens to more than one hundred. In addition a particular execution of a grasp configuration can be influenced by many unpre-

dictable factors. Consequently, each configuration is executed a sufficiently large number of times, by varying the location and orientation in the presentation of the object. In this way, statistically significant conclusions can be reached.

Nevertheless, this repetition could lead to a non practical number of executions, for each object only a few configurations are selected to be executed. This selection includes the most representative configurations of each object. Each configuration is executed 12 times: four times for three different orientations of the object.

## 4 Nonparametric prediction strategy

The data collected during the experiments comprises a large amount of information. Several analyses can be carried out over this data, especially those regarding the appropriateness and usefulness of the different features (see [2]). Here, however, we are more interested in the predictive capabilities that can be inferred from these data and the methods that can make the best use of it.

In theoretical terms a data set is composed of  $N$  executed grasps. Each grip  $g_i, i = 1 \dots N$  is described by the nine visual features  $q_{i1}, \dots, q_{i9}$  introduced in subsection 2.2. The space  $Q_S$  is formed by the ranges of the values of the features. Moreover, we have also recorded the performance of the grip and have assigned it to a class  $\omega_i \in \Omega$  for each  $g_i$ .

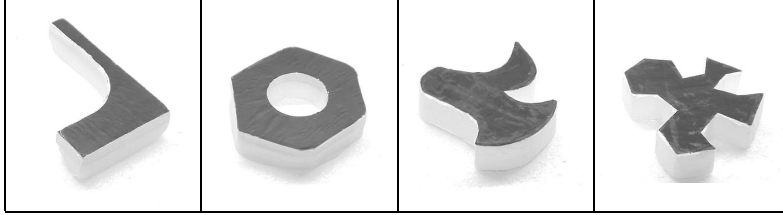
### 4.1 KNN classification rule

A prediction function has the form  $F(g) = \bar{\omega}$  where  $g \in Q_S$  and  $\bar{\omega} \in \Omega$ . There exists a wide bibliography on the building of such functions based on the Bayesian decision theory[5]. In this paper we have chosen the approach of the nonparametric techniques in particular the *voting k-nearest neighbors (KNN) rule* for modeling this function [11, 5]. The nonparametric techniques do not assume any density distribution of the features and the classes. To predict the class of a *query* point  $g_q$ , the KNN rule counts the K-nearest neighbors and choses the class that most often appears.

In our implementation we have introduced some modifications to the basic schema. First we use the Euclidean metric for measuring the distance between the points in  $Q_S$ . We weighted the contribution of each of the KNN points according to its distance to the query point. This gives more importance to the closer points. The kernel function used is  $K(d) = \frac{1}{1+(d/T)}$ , where  $T$  is an adjustable parameter, and  $d$  is the distance.

Given  $knn(g_q) = \{(g_i, \omega_i), i = 1 \dots k, g_i \in Q_S, \omega_i \in \Omega\}$  the  $k$  closest points to  $g_q$  and  $d_i$  is the corresponding distance from  $g_q$ . The votes or probabilities corresponding to a class  $\bar{\omega}$  are computed using this expression:

$$P(\bar{\omega}, g_q) = \sum_{\substack{g_i \in KNN(g_q) \\ \omega_i = \bar{\omega}}} \frac{K(d_i)}{\sum_{g_j \in KNN(g_q)} K(d_j)}$$



**Fig. 8.** The four objects used in the experiments.

Function  $P$  is also an expression of the posterior probability [11]. Our predictor would be defined as  $F(g_q) = \operatorname{argmax}_{\omega \in \Omega} P(\omega, g_q)$ .

#### 4.2 Error and risk functions

Performance of classification methods is measured in terms of successful or wrong classifications. Our classes have an important particularity, their qualitative order (e.g.: class A means a higher stability for a grip than any other class). Having this in mind, we try not to penalize in the same amount when the failure is qualitatively smaller (e.g.: predicting B when the outcome is C), than larger (i.e.: predicting A when the outcome is D). For this we build the error function  $E(\bar{\omega}, \omega)$ , being  $\bar{\omega}, \omega \in \Omega$ , where  $\bar{\omega}$  is the predicted outcome and  $\omega$  the real one. This is easily implemented with a table (see practical cases in Table 2).

A step further is the definition of the *risk function*:

$$R(\bar{\omega}, g_q) = \sum_{\omega \in \Omega} P(\bar{\omega}) E(\bar{\omega}, \omega), \text{ where } \bar{\omega} \in \Omega.$$

The class  $\omega \in \Omega$  selected for the prediction is the one that minimizes the risk,  $F(g_q) = \operatorname{argmin}_{\omega \in \Omega} R(\omega, g_q)$ . Using the risk function makes it possible to introduce in the prediction step the qualitative ordering of the problem classes.

#### 4.3 Implementation and results

A series of experiments were done following the experimental protocol described in section 3. Three different combinations of physical properties were tested: light objects and low friction (light/low), heavy objects and high friction (heavy/high); and light objects and high friction (light/high). A set of four different objects, depicted in Fig. 8 were used. Table 1 shows the number of different grips executed for each case, and the percentages of grips that resulted in each class of  $\Omega$ . Note that the total number of grips results from the repetition of a smaller number of configurations.

Two basic questions need to be answered about the prediction capabilities of the rules described in section 4.1: first, are they able to generalize across different objects, and second, did we have enough data to properly construct a

	E	D	C	B	A	Total
<b>Light</b>	102	84	33	27	18	264
<b>Low</b>	38.6%	31.8%	12.5%	10.2%	6.8%	(22)
<b>Light</b>	51	97	56	38	118	360
<b>High</b>	14.2%	26.9%	15.6%	10.6%	32.8%	(34)
<b>Heavy</b>	95	92	29	2	2	220
<b>High</b>	43.1%	41.8%	13.2%	0.9%	0.9%	(23)

**Table 1.** Sample distribution among classes for the different data sets. The figures in brackets in the “Total” column indicates the number of different configurations really tested.

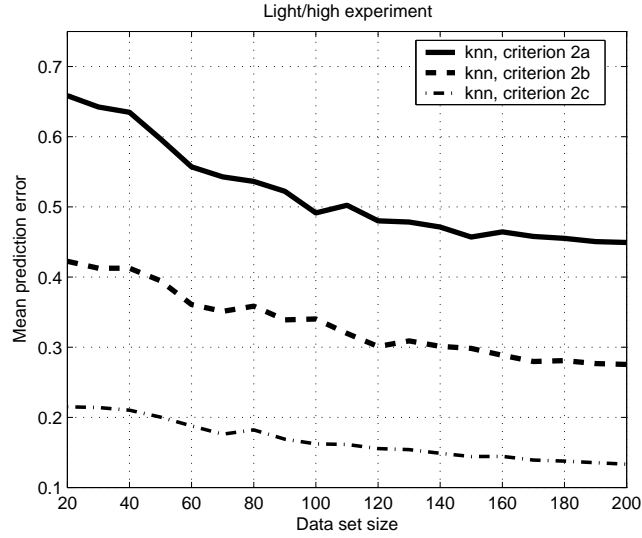
risk function? To answer these questions we have developed a cross-validation method named *leave-one-grasp-out validation* similar to the well known *leave-one-out validation* and *n-fold cross-validation* [5]. This consists of the following steps: 1) given the whole data set, remove all the points of a particular grasp configuration and use this subset as validation set; 2) use the remaining samples for predicting the outcomes of the validation set and compute the mean error; 3) repeat steps 1) and 2) for all configurations. The validation error will be the mean error of the iterations of step 2). The reason for removing all the points of a configuration from the data set is that all the points of a particular configuration are very close in  $Q_S$  and the KNN rule would be affected by this points instead of using points of unrelated configurations, farther in  $Q_S$ .

Moreover, we are also interested in the sensitivity of the error with respect to the size of the data set. We can analyze it by modifying the second step. Instead of using the whole remaining data set, we chose randomly a set of given size. This introduces a random factor, and to reduce the effect of this randomness we repeat this step a sufficiently large number of times

Criterion 2a						Criterion 2b						Criterion 2c					
	E	D	C	B	A		E	D	C	B	A		E	D	C	B	A
E	0.0	0.5	1.0	1.0	1.0	E	0.0	0.0	1.0	1.0	1.0	E	0.0	0.00	0.25	0.50	0.75
D	1.0	0.0	0.5	1.0	1.0	D	0.0	0.0	1.0	1.0	1.0	D	0.25	0.00	0.00	0.25	0.50
C	1.0	1.0	0.0	0.5	1.0	C	1.0	1.0	0.0	0.0	0.0	C	0.50	0.25	0.00	0.00	0.25
B	1.0	1.0	1.0	0.0	0.5	B	1.0	1.0	0.0	0.0	0.0	B	0.75	0.50	0.25	0.00	0.00
A	1.0	1.0	1.0	1.0	0.0	A	1.0	1.0	0.0	0.0	0.0	A	1.00	0.75	0.50	0.25	0.00

**Table 2.** Error tables. The rows indicate the predicted outcome, and the columns the real outcome. An error 1.0 indicates a failure, and 0.0 a successful prediction.

We have defined three error tables (see table 2). The first one *2a* is quite strict. It considers as failure any wrong prediction. The only exception is that it considers half a failure a prediction one class lower than the real output. This is a kind of conservative rule. The second table *2b* is a way of reducing the classes



**Fig. 9.** Size sensitivity validation for the data set of light objects and high friction

to two super-classes: the first one composed of class A, B and C, and the second, D and E. Finally, the third table  $2c$  tries to penalize errors depending on the qualitative distance between the predicted outcome and the real one. Again, “small” penalization is assigned for predicting a class as to be less stable than it really is.

The different parameters of the knn prediction rule,  $K$  and  $T$  for the kernel function, have been chosen using leave-one-out validation with the full datasets minimizing the errors.

Figure 3 shows the evolution of the prediction error for the data *light-high* set using the three error schemes. The first and most obvious observation that can be drawn from these figures is that the error is reduced as the size of the available data set increases. Moreover, the evolution of the error rates depending on the table error used seems to be equivalent, but in a different scale.

	Light/Low	Light/High	Heavy/high
<b>Criterion 2a</b>	0.438	0.449	0.245
<b>Criterion 2b</b>	0.236	0.275	0.115
<b>Criterion 2c</b>	0.110	0.133	0.035

**Table 3.** Error rates reached with the three data sets when the size of the data set is 200.

Finally Table 3 shows the error rates reached in the size sensitivity experiments with the three different data sets.

## 5 ANN classification approach

	Light/Low			Light/High		Heavy/High	
	ANN	ANN-2 <sup>2</sup>	knn	ANN	knn	ANN	knn
<b>Criterion 2a</b>	0.475	0.467	0.424	0.579	0.425	0.334	0.254
<b>Criterion 2b</b>	0.250	0.243	0.218	0.347	0.265	0.164	0.116
<b>Criterion 2c</b>	0.147	0.142	0.108	0.215	0.127	0.059	0.038

**Table 4.** Comparison between the error rates produced by the trained neural networks and the non-parametric classification approach. These results are obtained using the whole data sets as training sets.

A group of well known methods for classification are based on the use of artificial neural networks. In our approach to the problem we also have considered the use of this as a possible alternative.

For the neural network implementation we chose the classical feedforward backpropagation architecture. The net takes as inputs the values of the 9 features of a grip, and provide as output the choice between the 5 possible classes, classifying in this way the grip according to its expected reliability. The net has one hidden layer, made up by 14 nodes. The transfer function used in both middle and output layers is the logistic sigmoid. Bias inputs are added to both layers. The output layer is competitive, so that only one of the output goes to 1 whilst the others are forced to 0. The net parameters have been chosen through an experimental setting process. The networks were trained with the Rprop (resilient backpropagation) [17] and with the Levenberg-Marquardt [6] algorithms, obtaining slightly better results with the former.

In order to compare the results of the ANN approach with the non-parametric prediction approach we have applied again the *leave-one-grasp out validation*. In this case we have not performed an exhaustive sensitivity analysis of the size of the train data set. Neural-networks usually need large data sets to produce a good performance, so we have ours trained with the complete data sets. The only case in which we have studied the influence of the size is in the case of the light/low data set. In this case we disposed of a batch of 212 additional trials which were not classified in configurations. The results in tables 4 and 5 indicate that the performance of the neural network increases (from 73.5% to 78.4% of error type 0 and 1), with the additional data. Nevertheless, the performance of the knn prediction function is still slightly better (78.9%).

<sup>2</sup> The ANN-2 experiment is a special one. A set of 212 unclassified trials (which configuration is unknown), is added to the training set. The purpose is to check if there is any change in the results when the size of the training set is increased.

Error type	Light/Low				Light/High			Heavy/High		
	Random	ANN	ANN-2 <sup>2</sup>	knn	Random	ANN	knn	Random	ANN	knn
0	28.1%	39.0%	41.7%	38.7%	23.5%	35.9%	52.2%	37.9%	52.3%	63.6%
1	36.5%	34.5%	36.7%	40.2%	26.2%	28.3%	21.4%	47.4%	42.3%	33.2%
2	17.9%	13.2%	11.7%	15.5%	20.3%	22.0%	13.6%	12.4%	3.6%	2.3%
3	12.2%	10.3%	7.2%	4.9%	20.7%	12.2%	10.6%	1.5%	1.4%	0.9%
4	5.3%	3.0%	2.7%	0.7%	9.3%	1.6%	2.2%	0.8%	0.4%	0.0%

**Table 5.** Comparison of the composition of the error in the prediction. For each data set there are three columns. “Random” is the theoretical probability; “ANN” the error produced by the ANN prediction; and “knn” the error of the non-parametric classification using criterion 2a for the evaluation of the risk. The error type indicates the difference between the predicted class and the real outcome. Type 0 indicates a successful classification, type 1 indicates that the predicted class was one above or behind the real one.

Table 4 shows the error rates according to risk criteria defined in table 2 produced by both classification approaches. A first conclusion that can be learned from the tables is that the knn classification methods are better in all the cases. This conclusion is also supported by the results shown in tables 5 and 6. In these tables we also show the theoretical probability of choosing the class randomly and taking into account the probability of each class according to table 1. Table 6 shows a detailed description of the errors for one of the data sets. Again the knn classification shows better behavior than neural nets. Moreover both methods represent a neat improvement over the theoretical probability.

## 6 Discussion

From a practical point of view, when performing a stochastic action like grasping an unmodeled real object with a robotic hand, an error between two neighbor classes can be considered acceptable, especially in the case of a false negative. Indeed, it means that the reliability of the grasp is only slightly better than the predicted one. This justifies the definition of criteria 2b and 2c. The results summarized in table 3 suggest that the expected error rates will be around 0.25 (with error criterion 2b) or even close to 0.1 (with criterion 2c). It must be noted that these results have been obtained even though the available data were far from optimal. Firstly they were very unequally distributed across the classes, with some classes poorly represented. More precisely, the low-quality classes D and E strongly prevail on the others except in one of the data sets. Secondly, they were very noisy due to uncontrollable errors in sensing, image processing and motor control.

The comparison with the neural net approach yields the conclusion that the knn approach is clearly preferable. The reason is either we do not have enough data samples for training the net reliably (in spite of using the complete sets of



Random						ANN					
	E	D	C	B	A		E	D	C	B	A
E	2.0%	3.8%	2.2%	1.5%	4.7%	E	7.0%	5.9%	3.2%	0.9%	0.9%
D	3.8%	7.2%	4.2%	2.9%	8.8%	D	3.1%	11.1%	3.9%	3.9%	5.7%
C	2.2%	4.2%	2.4%	1.7%	5.1%	C	1.2%	3.2%	2.2%	3.2%	8.1%
B	1.5%	2.9%	1.7%	1.1%	3.9%	B	1.5%	1.2%	1.8%	1.8%	5.2%
A	4.7%	8.8%	5.1%	3.9%	10.8%	A	0.7%	4.1%	4.5%	2.0%	13.7%

Knn					
	E	D	C	B	A
E	8.6%	3.3%	1.4%	0.6%	0.8%
D	3.1%	14.7%	4.2%	2.8%	3.2%
C	1.1%	1.1%	0.6%	0.0%	0.3%
B	0.0%	0.8%	2.2%	1.1%	1.4%
A	1.4%	6.9%	7.2%	6.1%	27.2%

**Table 6.** Detailed error tables of the values summarized in table 5 for the data set “Light/High”. The rows indicate the predicted class for query, while the columns indicate the real class of the query. Each cell of a table contains the percentage of predictions of class *row* when the real output was the class *column*.

data points), or, simply, that the neural net, or our particular implementation, is not well suited to this problem.

In addition, the knn classification has shown several interesting properties. First, with a reasonably small number of samples it is able to produce quite good classification rates. Table 5 shows that in all the data sets the rate of error with distance of type 1 or 0 (i.e.: successful) is always over 70%, reaching 96.8% in the “Heavy/High” data set. This significant figure in such a stochastic problem like this. A second conclusion is that error rates decrease as the size of the data set increases; and a good level of error rate is achieved with less than 200 training samples. Third and final, the use of the *leave-one-grasp-out validation* technique for measuring the errors allows us to conclude that generalization is possible within this scheme, so the output of a given configuration can be predicted from the experience of different but similar configurations, for the same object or for a different object.

## 7 Conclusions

We have presented a contribution to a methodology for computing and executing reliable grasps for unmodeled objects using only visual sensing as input, in such a way that the system can exhibit an incremental adaptive behavior based on its previous experiences. We have proposed a set of intrinsic features that adequately characterize a grip and can be computed by using only the object image and the kinematics of the hand. We have implemented a number of different approaches that accept as inputs the features, and give as output the reliability class of the

grip. Feature space data were obtained from real experiments with a humanoid robot. We used both features that depend on the object image and also on the particular kinematic configuration of the hand.

The obtained prediction results are satisfactory enough to suggest that the methodology is adequate and further progress should be made in this direction. This progress should follow the line of designing a learning strategy that minimize the number of experimental trials necessary to build the data set used for the prediction of successful future trials.

## References

1. A. Bicchi and V. Kumar. Robotic grasping configuration and contact: A review. In *IEEE Intl Conf. on Robotics and Automation*, April 2000.
2. E. Chinellato. Robust strategies for selecting vision-based planar grasps of unknown objects with a three-finger hand. Master's thesis, School of Artificial Intelligence. Division of Informatics. University of Edinburgh, 2002.
3. E. Chinellato, R.B Fischer, A. Morales, and A.P. del Pobil. Ranking planar grasp configurations for a three-finger hand. *IEEE Intl. Conf. on Robotics and Automation*, 2003.
4. J.A. Coelho Jr. and R.A. Grupen. Online grasp synthesis. In *Proc. 1996 IEEE Int. Conf. Robotics Automat.*, volume 3, pages 2137–2142, Minneapolis, Minnesota, April 1996.
5. R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2nd edition, 2001.
6. M.T. Hagan and M. Menhaj. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993, 1994.
7. Barrett Technology Inc. <http://www.barrett.com/>.
8. M. Jeannerod, M.A. Arbib, G. Rizzolatti, and H. Sakata. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neuroscience*, 1995.
9. A.D. Milner and M.A. Goodale. *The visual brain in action*. Oxford University Press, 1995.
10. B. Mirtich and J. Canny. Easily computable optimum grasps in 2d and 3d. In *IEEE Intl Conf. on Robotics and Automation*, pages 739–747, May 1994.
11. Tom M. Mitchell. *Machine Learning*. McGraw Hill, 1997.
12. D.J. Montana. The condition for contact grasp stability. In *IEEE Intl. Conf. on Robotics and Automation*, Sacramento, California, 1991.
13. A. Morales, P.J. Sanz, and A.P. del Pobil. Heuristic vision-based computation of three-finger grasps on unknown planar objects. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pages 1693–1698, Lausanne, Switzerland, 2002.
14. A. Morales, P.J. Sanz, A.P. del Pobil, and A. H. Fagg. An experiment in constraining vision-based finger contact selection with gripper geometry. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pages 1711–1716, Lausanne, Switzerland, 2002.
15. Y.C. Park and G.P. Starr. Grasp synthesis of polygonal objects using a three-fingered robot hand. *International Journal of Robotics Research*, 11(3):163–184, 1992.
16. R. Platt, A.H. Fagg, and R. Gruppen. Nullspace composition of control laws for grasping. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pages 1717–1723, Lausanne, Switzerland, 2002.

17. M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *IEEE Intl. Conf on Neural Networks*, 1993.
18. G. Rizzolatti and G. Luppino. The cortical motor system. *Neuron*, 31:889–901, 2001.
19. H. Sakata, M. Taira, M. Kusunoki, A. Murata, and Y. Tanaka. The parietal association cortex in depth perception and visual control of the hand action. *Trends in Neuroscience*, 1997.