VMGdB
the CONTACT Visuo Motor Grasping dataBase

N. Noceti   C. Castellini   B. Caputo   G. Sandini
F. Odone

October 1, 2009

Abstract

Grasping is one of the most interesting challenges in nowadays robotics, posing problems to the mechanical and electronic engineer, the computer vision researcher, the control theorist and, more recently, the neuroscientist. The study of human grasping has proved beneficial to get a better understanding of the problem. In this paper we present VMGdB, the CONTACT Visuo-Motor Grasping Database, a recording of grasping actions performed by 20 human subjects on 7 objects using 5 ways of grasping, under variable illumination conditions. The VMGdB consists of 5200 grasping acts organized in 260 data entries — each of which made of 2 video sequences recorded from two colour cameras, and motor data recorded from a sensorised glove. Labeled data are available as standard AVI videos and a file of ASCII outputs from the glove. The VMGdB provides to the community a reliable and flexible testbed for tackling the problem of grasping from a humanoid/human-oriented perspective, and hopefully not only that.

1 Introduction

Grasping is a serious challenge to nowadays robotics. As research moves towards more and more unstructured environments, the characteristics of the objects to be grasped (shape, texture, weight) become unpredictable; the illumination and colour conditions of the scene must be taken into account; and, as better and better grippers/hands are available, the kinematics and dynamics of the end-effector must enter the picture, too. As an example, consider a rover equipped with a hand-arm system intended to pick up samples of Martian rocks and soil, and to then drop them in the correct test tube: such a system must be able to grasp with a so-far unprecedented ability.

A recent trend to solve this problem (to which so far no general solution is known) draws inspiration from human grasping and its developmental nature. This is in the first place inspired by the amazing ability of humans to grasp; for instance Gibson's re-definition of objects in terms of affordances [Gib77, Gib86] entails that the way an object can be grasped is essential to its model, which can then be used in a robotic artifact. More recently, this idea has been boosted by the discovery of a neural correlate to affordances, namely mirror neurons and mirror structures, both in high primates and humans [GFFR96, UKG^{+}01, RC04, KNW^{+}09]. If this paradigm is correct, then studying human grasping is paramount for good robotic grasping.

Starting from this working hypothesis, we hereby present to the community VMGdB, the CONTACT Visuo-Motor dataBase, obtained by recording (both in video and with a dataglove) 5200 grasping acts, performed by 20 human subjects on 7 different objects and with 5 different grasp types.

1.1 Motivation and aims

The intentionally unstructured illumination conditions, and the fact that the objects are of the most diverse shapes, textures and colours, make this database a rather realistic ground model of the human act of grasping. Two different points of view and the use of a magnetic tracker plus a dataglove, ensure that we have a clear representation of the action. The fact that the VMGdB collects such diverse data as far as the subjects/objects/grasps are concerned makes it suited, for instance,

- to model and understand the act of grasping itself (prediction of the hand final position and posture, detection of the preshaping phase, classification of the type of grasp, etc.);
- to enrich the visual model of objects, adding the knowledge of the type of grasp employed for each object;
- if a task-related prior is given, to then choose an appropriate robotic grasp; and so on.

Even though grasping is being extensively studied, as far as we know this is the first publicly available database aimed at the study of grasping as a dynamic act. Interesting similar, but not comparable, attempts appear in the GRASP project [Kra09], where a highly detailed taxonomy of grasp types is described, and in the Columbia Grasp Database [GCDA09], which consists of synthetic grasps generated using a simulator.

The VMGdB follows the idea described in [LSV05, MSN^{+}06] and it represents an extension and an advancement of it, both in a qualitative and quantitative sense.

1.2 Paper structure

The paper is organized as follows. Section 2 details the acquisition protocol, describing the how the acquisition setup and the experiments were designed, Section 3 illustrates the dataset and additional material included to the VMGdB, and Section 4 is left to the conclusions.

2 The acquisition protocol

This section sketches the main elements of the acquisition system and details on the acquisition experimental procedure.

2.1 Setup

The acquisition setup was designed in order to give an accurate and meaningful representation of the act of grasping an object by the volunteers. Its main components are two colour cameras for the visual representation; and a magnetic tracker, a virtual-reality sensorised glove and a pressure sensor for the haptic/motor representation. All devices are connected to a standard biprocessor desktop computer equipped with a hard disk large enough to contain all required data, and fast enough not to loose any data while recording.

Human subjects would sit on a comfortable chair in front of a desk onto which an object would be placed; the subject was asked to grasp the object with his/her right hand while wearing the glove, and then to put it back in the original position on the table with the left hand, as the right hand goes back to the resting position. Therefore, a full grasping act goes from the hand resting position, to the grasping instant and then back to the resting position.

The desk is uniformly dark green and non-reflective; the objects were chosen to be colourful; and the illumination was provided by two windows looming over the desk. Intentionally we did not fix the illumination condition (that changed over time since acquisition sessions spanned over a week), now and then fixing the white balance of the cameras in order to avoid saturation.

The cameras are Watec *WAT-202D* colour cameras, operating at 25Hz and connected to two Picolo PCI-bus frame grabbers. One camera was placed in front of the subject while the other was placed on the right-hand side of the subject, almost framing the object in a close-up and focussed upon it. The first camera has the view of what an external observer would be seeing of the grasp; the second would give an accurate representation of the act of grasping in full detail, including the last moments of the reaching sequence.

The subjects would wear a 22-sensors Immersion *CyberGlove* [Vir98] right-hand sided dataglove, which provides 22 8-bit numbers linearly related to the angles of the subject's hand joints. The resolution of the sensors is on average about 0.5 degree. The sensors describe the position of the three phalanxes of each finger (for the thumb, rotation and two phalanxes), the four finger-to-finger abductions, the palm arch, the wrist pitch and the wrist yaw. The magnetic tracker was an Ascension *Flock-Of-Birds* [Asc99] mounted on the subject's wrist, which would return six real numbers, the linear and angular coordinates of the wrist with respect to a base mounted on the far end of the desk. Lastly, a standard force sensing resistor (FSR) glued to the subject's thumb was used to determine the instant of contact with the object.

Table 1: The 13 *(grasp,object)* pairs in the VMGdB. Each grasp is performed on the related object 20 times by each of the 20 subjects, resulting in 5200 grasping acts. Boldface **X**s are the pairs visibile in Figure 1, bottom row.

	ball	pen	duck	pig	hammer	tape	lego brick
cylindric power				**X**			
flat					**X**		X
pinch		**X**	X			X	X
spherical	**X**					X	
tripodal	X	X	**X**			X	

2.2 Experiment design

The dataset is built considering 7 different objects ((see Fig. 1), top) and 5 grasps ((see Fig. 1), bottom). The objects have been selected to represent different materials, colors and shapes, and to them we associate a set of appropriate grasps [Kra09]: each object can be grasped, in general, in many different ways, according to the many-to-many relationship reported in Table 2.2. In total 13 *(grasp,object)* pairs are considered, according to everyday experience.

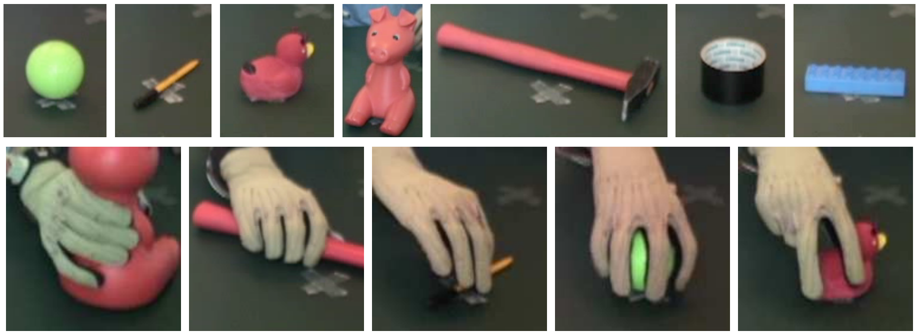

Figure 1: *(top row)* The 7 objects in the VMGdB. *(bottom row)* The 5 grasp types in the VMGdB (left to right: cylindric power grasp, flat, pinch grip, spherical and tripodal grip) applied to five of the considered objects.

The subjects pool includes 20 right-handed people, 6 females and 14 males, aged between 24 and 42 years (mean 31.5 years, median 31). Each of the 260 *(subject,object,grasp)* triplets is an entry of the database; each entry was performed 20 times, giving a total of 5200 grasping acts.

The next section describes in details the available data for each experiment.

3 The VMGdB

In this section we report the structure of the VMGdB, with a comment on the peculiarities of the data and a brief description on the data annotation features.

Table 2: Items contained in each row of the sensor data file.

#	data type	comments
1	real	time stamp (starting from 0)
6	real	measurements from the Flock-of-Birds
22	8-bit integer	measurements from the Cyberglove
1	$16 - bit$ integer	measurement from the pressure sensor

3.1 Data

Each of the 260 *(subject, object, grasp)* entries is associated to the following data:

- **Visual information.** Two video sequences (384×288 25 fps AVI, MPEG-4 compression, YV12 colorspace) acquired by the two cameras with a focus on the object and on the action respectively. The videos report the whole grasping action, from a hand rest position to the final grasp and then back. Figure 2 shows sample frames from the two sequences. Each video sequence is associated to a data file (ASCII format) mapping single video frames to the acquisition time-stamp, allowing for synchronization with the sensor data.

- **Hand posture sensor information.** One file (ASCII format) containing information from the sensor glove as the grasping action occurs as well as the acquisition time-stamp (Table 3.1 lists in details the content of each row of the file). The latter is used for synchronization with the visual data.

Both videos and ASCII files contain all information related to a whole *(subject, object, grasp)* experiment, i.e., to 20 consecutive grasping actions performed by the same subject on a given object-grasp pair.

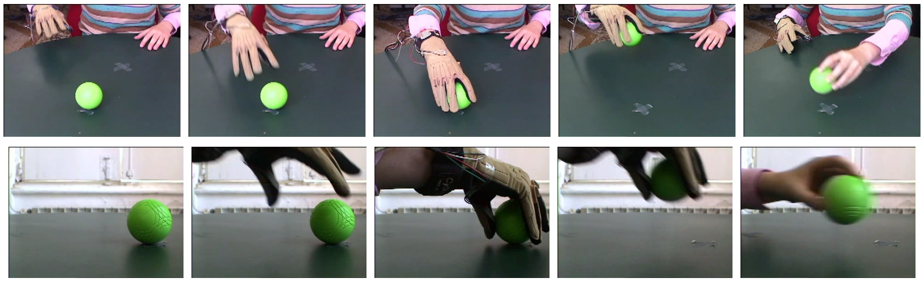

Figure 2: Synchronized sample frames from the two video sequences, showing the evolution of the grasping action (from the hand rest position to object lifting and then back to rest).

3.2 Additional material

Together with the data described above, supplemental material is made available to increase the database usability.

Object visual appearance. A set of 20 plain images of each object depicted in the acquisition environment is included as additional material. A crop of these images to contain the sole object region is also available.

Synchronization. The two video sequences have been acquired at 25Hz by each camera, while the glove, magnetic tracker and pressure sensor have been sampled at 100Hz. Since the three devices are independent of one another, a time-stamp is associated to each numerical sample and frame (for each of the cameras) in order to reconstruct the exact time sequence of data and to obtain a full synchronization of visual and sensorial information. A script associating each sensor datum the corresponding video frames is available; because of the different acquisition frequency of the different devices there will not be, in general, an exact correspondence between time-stamps. Thus the script implements a nearest neighbor strategy to find the video-frames closest in time to the sensor datum.

Istantaneous visuo-motor data. The database contains information on the whole grasp action, from the hand rest position and back. This can be useful to researchers wishing to exploit the action evolution for what concerns visual or motion features, or both. In the case instantaneous information on the actual grasping is needed we included a script which extracts static visual and motion features from the dynamic data sequences. When applied to a (*subject, object, grasp*) experiment, it extracts the 20 image frames and an ASCII file with the 20 vector measurements of the joints corresponding to the 20 grasps within this experiment.

The grasping instant is estimated by locating extrema from the pressure sensor. Figure 3 reports the values of the pressure sensor positioned on the thumb (first of the 4 pressure values), along an experiment. The zoomed plot shows in details a single action from one rest position to the next rest position.

3.3 Data peculiarities

The data have been acquired under loosely controlled conditions. As mentioned in Section 2.1, during acquisition the scene was illuminated by natural light with a significant illumination change from sequence to sequence.

Also, the subjects were encouraged to act "naturally", therefore there is a certain variability in the way actions were performed. For most subjects, it is also noticeable a speed increase as the experiments proceed and the actor gets familiar with the experimental protocol.

All these features make the visual data very close to natural acquisition settings and give raise to typical issues of an artificial vision system — non uniform illumination, shadows, occlusions, just to name a few (see Figure 4). For these reasons the VMGdB is an ideal test-bed for validating solutions to be then applied to the real world.

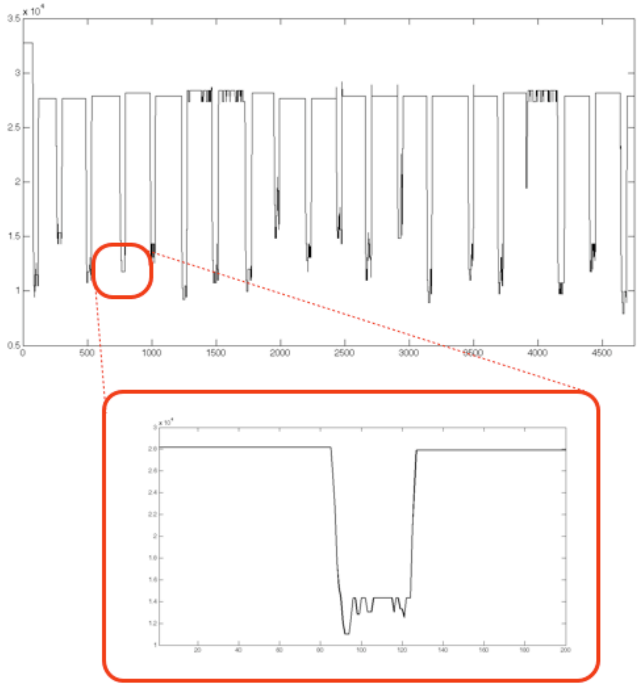

Figure 3: A plot of the thumb pressure sensor values used to estimate the 20 grasping instants within one experiments. The zoom shows in details a single action.

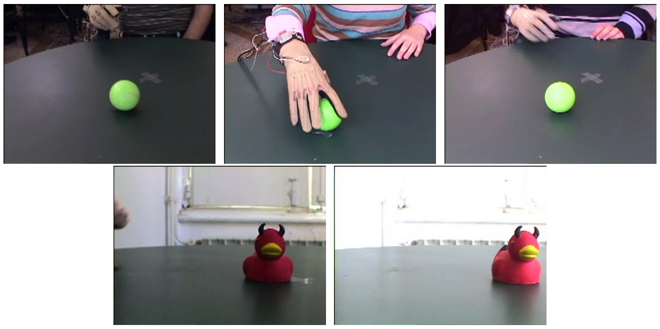

Figure 4: 3 frontal (top) and 2 lateral (bottom) sample frames of the same nature extracted from different experiments, showing the variability of illumination conditions and, consequently, the varying amount of shadows.

4 Conclusions

The CONTACT Visuo-Motor Grasping Database, VMGdB, has been presented in this paper. It consists of a large collection of synchronised video sequences and sensor data acquired by a sensorized glove, faithfully representing the act of human grasping.

The data was obtained with the help of 20 human subjects engaged in grasping 7 common objects in 5 different ways. In total, there are 260 experiment entries, recording 5200 grasping actions. The acquisition setting was intentionally loosely controlled, thus visual appearance may change due to variable illumination, shadows and occlusions caused by the actors hands. For these reasons the VMGdB represents a rather challenging set of data, ideal to depict common problems of an artificial vision environment.

Along with visuo-motor dynamic data, recording the whole grasping sequence, the VMGdB includes images of the plain objects and a script to extract instantaneous static information (both visual and sensorial) on the precise instant when grasping occurs. This feature would make the database useful to researcher not wishing to exploit the full grasping dynamics.

The VMGdB is publicly available at ... **CLAUDIO SAYS: qui dobbiamo decidere COME renderlo pubblico. FRA: proporrei una semplice pagina web con materiale da downloadare, un readme file semplice e questo papiro, quando ci sara' piu' eventuali lavori correlati. Una cosa cosi': http://www.bioid.com/downloads/facedb/index.php o anche piu' semplice. Sul "dove" il nostro sito e' attualmente in fase di aggiornamento ma nel caso saremo felici di ospitare il db. Altrimenti IDIAP? CC: non posso che approvare!**.

Acknowledgements

The authors are grateful to the twenty volunteers that participated to the acquisition sessions. The work here described was partially supported by the EU-funded projects NEURObotics (FP6-IST-1917) and Contact (NEST-5010). We also thank Giorgio Metta and Giulio Sandini of IIT, Genova, Italy, for an early discussion about the experiment this database is based upon.

References

[Asc99] Ascension Technology Corporation, PO Box 527, Burlington (VT), USA. *The Flock of Birds — Installation and operation guide*, January 1999.

[GCDA09] Corey Goldfeder, Matei Ciocarlie, Hao Dang, and Peter K. Allen. The columbia grasp database. In *proceedings of the IEEE International Conference on Robotics and Automation*, 2009.

[GFFR96] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119:593–609, 1996.

[Gib77] James J. Gibson. *The Theory of Affordances*, chapter 8. Erlbaum Associates, 1977.

[Gib86] James J. Gibson. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1986.

[KNW$^+$09] J. M. Kilner, A. Neal, N. Weiskopf, K. J. Friston, and C. D. Frith. Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci.*, 29:10153–10159, 2009.

[Kra09] Danica Kragic. GRASP: Emergence of cognitive grasping through introspection, emulation and surprise, 2009. Funded by the European Commission under code IST-FP7-IP-215821.

[LSV05] M. Lopes and J. Santos-Victor. Visual learning by imitation with motor representations. *IEEE Transactions on Systems, Man, and Cybernetics, Part B Cybernetics*, 35:438–449, 2005.

[MSN$^+$06] G. Metta, G. Sandini, L. Natale, L. Craighero, and L. Fadiga. Understanding mirror neurons: a bio-robotic approach. *Interaction Studies*, 7:197–232, 2006.

[RC04] G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192, 2004.

[UKG$^+$01] M.A. Umiltá, E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, and G. Rizzolatti. I know what you are doing: A neurophysiological study. *Neuron*, 31:1–20, 2001.

[Vir98] Virtual Technologies, Inc., 2175 Park Blvd., Palo Alto (CA), USA. *CyberGlove Reference Manual*, August 1998.