

# Human Grasping Interaction Capture and Analysis

Benjamin Verider  
McGill University  
benjamin.verdier@mail.mcgill.ca

Sheldon Andrews  
École de Technologie Supérieure  
sheldon.andrews@etsmtl.ca

Paul G. Kry  
McGill University  
kry@cs.mcgill.ca

## ABSTRACT

We design a system to capture, clean, and segment a high quality database of hand based grasping and manipulation. We capture interactions with a large collection of everyday objects. Optical marker-based motion capture and glove data are combined in a physics-based filter to improve the quality of thumb motion. Sensors stitched into our glove provide recordings of the pressure image across the fingers and the palm. We evaluate different segmentation techniques for processing motion and pressure data. Finally, we describe examples that explain how the data will be useful in applications such as virtual reality and the design of physics-based control of virtual and robotic hands.

## CCS CONCEPTS

• **Computing methodologies** → **Motion capture; Motion processing;**

## KEYWORDS

hands, grasping, interaction capture, segmentation

### ACM Reference format:

Benjamin Verider, Sheldon Andrews, and Paul G. Kry. 2017. Human Grasping Interaction Capture and Analysis. In *Proceedings of SCA '17, Los Angeles, CA, USA, July 28-30, 2017*, 2 pages.  
DOI: 10.1145/3099564.3108163

## 1 INTRODUCTION

How can we best capture human hands interacting with objects? This is a question we have been asking ourselves for some time. But the question must be asked in the context of the goals and the applications, which in our work is the capture of interaction strategies.

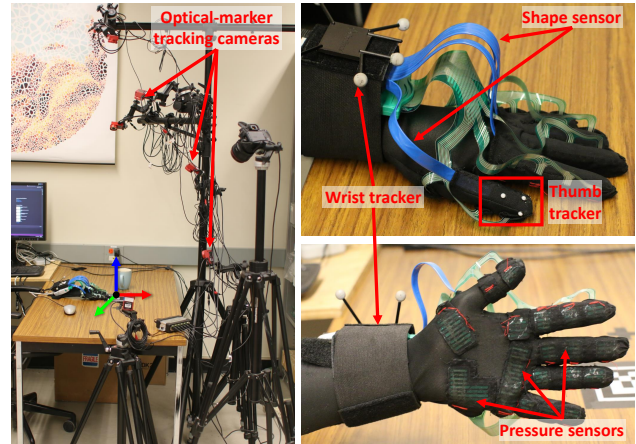
Many challenges come up when one tries to capture hand motions. For example, occlusion is a problem if only an optical tracking system is used. The use of a data glove resolves these, but may impede the motions of the wearer. We accept these limitations as the system allows a greater ease in capturing a large amount of data. However, given the poor quality of the thumb motion, we correct it using optical motion capture of the tip of the thumb, which is also used to record the rigid motion of the forearm.

Along with hand motions, we also capture the pressure across the fingers and the palm, extending Kry and Pai [2006]. We use this

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SCA '17, Los Angeles, CA, USA

© 2017 Copyright held by the owner/author(s). 978-1-4503-5091-4/17/07...\$15.00  
DOI: 10.1145/3099564.3108163



**Figure 1: Our capture setup. The capture volume with its origin (left); a close-up of the individual sensors that make-up the glove sensors (right, top and bottom). The glove combines shape and pressure sensors to capture both the hand posture and interaction forces.**

data in three different applications: segmentation, classification, reconstruction.

We believe that the capture and analysis of pressure data, along with hand motion, is an important first step toward capturing interaction strategies from which dexterous controllers can be built.

## 2 DATA COLLECTION

The hardware used in our capture setup is shown in Figure 1. The ShapeHand<sup>1</sup> is a glove with shape sensors running along each finger and the thumb. Its software outputs 16 quaternions, effectively describing the orientation of the each phalange and the wrist. Pressure data is collected using the Grip System by Tekscan<sup>2</sup>, which is composed of 361 tactile sensing cells distributed along the fingers and the palm. Each capture session is also monitored using an RGB camera. This provides a snapshot of the ground truth motion and allows for easy visualization of the grasps and manipulations in a human viewable format. All devices are software synchronized and sampled at a rate of 50 Hz. Each frame is also timestamped.

A rigid tracking cluster, consisting of several optical markers, is attached near the wrist and tracks the forearm motion. Another smaller tracking cluster is located near the thumb tip and is used to correct the thumb motion.

Our dataset contains a total of 211 captured sequences involving 50 objects. Each sequence involves grasping and manipulation tasks using the left hand. The set of objects is chosen to be diverse and includes kitchen utensils, tools, mugs, jars, and toys.

<sup>1</sup>Measurand ShapeHand. <http://www.shapehand.com/>

<sup>2</sup>Tekscan Grip. <http://www.tekscan.com/>

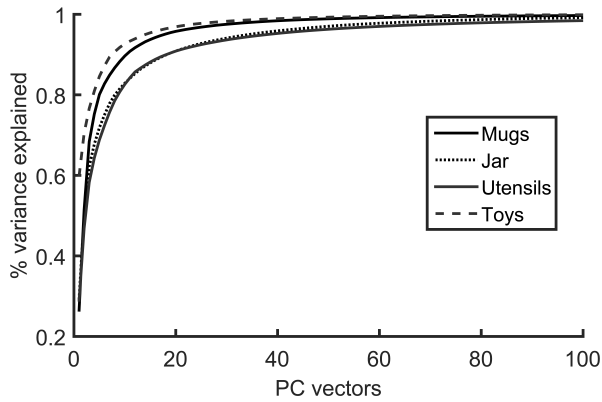


Figure 2: Variance in grasping and manipulation data is explained by only a small number of synergies.

### 3 DATA ANALYSIS

In this section we describe several example applications of our dataset including segmentation, classification, and reconstruction. We perform segmentation of the grasping interactions by an adaptation of a technique previously used only to segment full-body motions. Grasp classification using various combinations of features in our dataset is also performed. Finally, a machine learning approach is used to reconstruct hand postures just from pressure data. We begin with an analysis to motivate the reader that human grasping interactions are well described by low-dimensional embeddings.

#### 3.1 Low-dimensional grasping

Inspired by Santello et al. [1998], we perform a statistical analysis of grasping interactions for some of the objects in our set. The combined pose and pressure data provides rich information about the manipulation tasks. However, there is clearly a synergy of pressure and joint motion being used to manipulate the objects. Although, the dimensionality of the interaction data is high (361 pressure values and 64 quaternion components), 80% of variance in the data can be explained by fewer than 10 principal component (PC) vectors (see Figure 2). This coordination also appears when examining only the pressure data.

#### 3.2 Segmentation

From Barbič et al. [2004], we decide to adapt the method using PCA to be applicable to our dataset, since we have seen that a low-error projection of grasping information in a lower-dimension space is possible. This segmentation method is also straight-forward to implement and less complex than the probabilistic PCA method.

A tendency we expected and that we observe is that when using different types of data (shape, pressure) to perform the segmentation on the same sequence of capture, we obtain cuts that are close to each other but rarely at the same frame. The order is however coherent: the earliest cut is often determined by the pressure, the latest by the shape. However, we sometimes observe the inverse phenomenon, most of the time when grasping an object rather than releasing. This is due to the pre-shaping of the hand in anticipation for the future grasp. In this case, the joint's orientations become

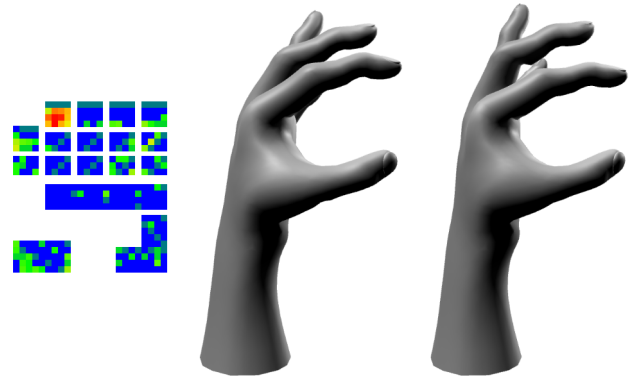


Figure 3: Example of pose reconstructed from pressure data. This shows the captured pose (center), the pose predicted by our trained CNN (right) and the corresponding pressure image (left)

close to their final position before any actual contact between the hand and the object.

#### 3.3 Identification and Reconstruction

After isolating the joints and pressure data corresponding to the type of grasps we captured, our next goal is to identify each grasp from the raw data, as well as reconstructing joint orientation from pressure data and *vice versa*. Our preliminary approach used a classifier based on PCA, since, given what we observed during the segmentation process, the dimensionality of the grasping data seemed to be easily reducible without too much loss. We generated a space of lower dimension from our data and projected each type of grasp. However, using PCA proved to be insufficient for effectively differentiating sets of distinct grasps. Our second idea was to use a convolutional neural network (CNN) in order to identify different grasps, as well as predict hand motions, only from pressure data. The identification through neural networks is highly successful. Processing the testing set with the trained neural network results in 98% of accuracy in identifying the grasps, and 92% regarding the object/grasp pairs. The pose prediction yields outstanding results as well.

### 4 FUTURE WORK AND CONCLUSION

We present a unique ensemble of sensors for capturing rich interactions involving human grasping. Our dataset includes hand poses and pressure information as well as large scale forearm and object motion. We provide an analysis of data collected using our system to motivate its use in computer graphics and virtual applications. The next step is to integrate this type of data into a physically based character controller, for instance.

### REFERENCES

- Jernej Barbič, Alla Safonova, Jia-Yu Pan, Christos Faloutsos, Jessica K Hodgins, and Nancy S Pollard. 2004. Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004*. Canadian Human-Computer Communications Society, 185–194.
- Paul G Kry and Dinesh K Pai. 2006. Interaction capture and synthesis. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 872–880.
- Marco Santello, Martha Flanders, and John F Soechting. 1998. Postural hand synergies for tool use. *Journal of Neuroscience* 18, 23 (1998), 10105–10115.