Robotics and Autonomous Systems 74 (2015) 108–121

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Recognizing the grasp intention from human demonstration

Ravin de Souza [a,b,*], Sahar El-Khoury [a], José Santos-Victor [b], Aude Billard [a]

[a] Learning Algorithms and Systems Laboratory, École Polytechnique Fédéral de Lausanne, Switzerland
[b] VisLab, Institute for Systems and Robotics, Insituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

HIGHLIGHTS

- We propose a general method to interpret human grasp behavior in terms of opposition primitives.
- A primitive model consisting of 41 oppositions for the hand is defined.
- The most likely primitive combination is inferred from tactile and configuration data.
- An 87% recognition rate is achieved over a wide range of human grasp behavior.

ARTICLE INFO

Article history:
Received 6 August 2014
Received in revised form
2 July 2015
Accepted 10 July 2015
Available online 17 July 2015

Keywords:
Grasp representation
Grasp recognition
Learning from demonstration
Tactile sensing

ABSTRACT

In human grasping, choices are made on the use of hand-parts even before a grasp is realized. The human associates these choices with end-functionality and is confident that the resulting grasp will be able to meet task requirements. We refer to these choices on the use of hand-parts underlying grasp formation as the *grasp intention*. Modeling the grasp intention offers a paradigm whereby decisions underlying grasp formation may be related to the functional properties of the realized grasp in terms of quantities which may be sensed/recognized or controlled. In this paper we model grasp intention as mix of oppositions between hand parts. Sub-parts of the hand acting in opposition to each other are viewed as a basis from which grasps are formed. We compute a set of such possible oppositions and determine the most likely combination from the raw information present in a demonstrated grasp. An intermediate representation of raw sensor data exposes interactions between elementary grasping surfaces. From this, the most likely combination of oppositions is inferred. Grasping experiments with humans show that the proposed approach is robust enough to correctly capture the intention in demonstrated grasps across a wide range of hand functionality.

1. Introduction

Significant research has been conducted into the design and development of anthropomorphic hands to enable robotic systems [1] as well as prosthetic devices [2] to interact with real-world environments. However the additional flexibility associated with these hands comes at the cost of complexity in control due to the increased degrees of freedom. While these hands can be tele-operated to perform complex grasping and manipulation tasks, autonomous control in response to the demands of a task scenario remains a difficult problem. Humans however are extremely adept at controlling the high degree of freedom hand–wrist–arm musculo-skeletal system and are able to grasp and manipulate objects according to task requirements with a minimum of effort. It is of interest therefore to study human grasping behavior in order to extract underlying principles which may be transferred to robotic or prosthetic devices [3–8]. This paper focuses on an aspect of human grasp behavior which we will refer to as the *grasp intention*.

Even before a grasp is realized, choices have been made regarding parts of the hand that will be engaged and the manner of their application against object surfaces. These choices stem from a perception of task demands and are therefore related to functionality that is brought to the grasp in terms of generating and controlling force, torque and motion. This is evident from the four task scenarios shown in Fig. 1. Tasks requiring dexterity (turning a dial, writing), make use of the finger tips which open up degrees of freedom and bring into play required manipulability for in-hand motion. Also, greater sensitivity associated with finger

---

[*] Corresponding author at: Learning Algorithms and Systems Laboratory, École Polytechnique Fédéral de Lausanne, Switzerland.

E-mail addresses: ravin.desouza@epfl.ch (R. de Souza), sahar.elkhoury@epfl.ch (S. El-Khoury), jasv@isr.ist.utl.pt (J. Santos-Victor), aude.billard@epfl.ch (A. Billard).

http://dx.doi.org/10.1016/j.robot.2015.07.006
0921-8890/

Fig. 1. Four task scenarios: turning a dial, writing, opening a tightly closed bottle-cap, screw-driving show the different use of hand-parts to provide grasp functionality as demanded by a task.

tips is essential for controlling the manipulation [9]. In contrast, tasks requiring power (opening a tight bottle cap, screw-driving) make use of finger surfaces and the palm. Use of these hand-parts is directly related to transmission of torque and motion generated by the wrist–arm system.

We refer to these choices on the use of hand-parts underlying grasp formation as the *grasp intention*. Grasp intention guides the realization of a stable grasp, determining outwardly visible posture and also the grasp function or the particular quality in which force, torque and motion can be delivered to the grasped object. Thus, modeling grasp intention provides a way to relate the decisions underlying grasp formation to end-functionality. This has applications in constructing an appropriate grasp for a task. From a learning from demonstration perspective, we are better positioned to decide on the more important aspects of a grasp to transfer to the robotic platform under consideration.

Is it possible to characterize grasp intention in some general way and recognize this from a grasp demonstration? Grasp taxonomies such as [3,10], based on studies of human grasp behavior, have been proposed as a means to capture grasp ability. The authors in [11–13] attempt to recognize a taxonomy category from human demonstration using cues such as visual features of the grasp or joint angles from a data glove. In [14,15] tactile information is incorporated as well. While it is useful to identify a taxonomy category, key information is lacking on how to recreate the grasp or adapt it to a different object while preserving underlying functional roles of the fingers involved. For example if the object is perturbed or used in a task context, are all hand surfaces equally important for applying pressure or are some more important than others. Similarly, if the properties of the object change how can we purposefully change the hand configuration and object contacts made while remaining confident that the essential meaning of the grasp is preserved. Heuristics have to be designed on a case by case basis to encode the meaning of each grasp. A more general approach defines a set of grasp components from which a wide range of grasps may be constructed. The problem is then identifying and prioritizing the appropriate set of components present in a grasp demonstration.

In this paper we adopt the hypothesis of Iberall et al. in [4] that opposition between hand-parts, while engaging the hand in a well-defined manner, is also correlated with the end-function to be delivered on a grasped object. Thus it is well suited to model grasp intention. Accordingly, a grasp is interpreted as a mix of oppositions between hand-parts. Each opposition serves a particular functional end. For example, the grasp of screw-driving in Fig. 2 may be interpreted as a combination of 3 components: action of the thumb against side of the fingers which supports the action of fingers against the palm in order to keep the tool gripped firmly, while use of the thumb-tip against the finger-tip enables the tool to be directed appropriately during the task. We infer this mix of oppositions from the hand configuration and tactile information in a grasp demonstration.

Although the Opposition Space model admits oppositions where the hand is working against external forces, this paper relies only on opposition between two hand-parts. Consequently, we are not able to recognize non-prehensile grasps, such as the hook or flat-palm grasps, which work against gravity. Similarly, with prehensile grasps, components where hand-parts work solely against task forces cannot be recognized. Examples of these would be finger extension for applying cutting force, pressing a button or resting the side of the palm against a surface during writing.

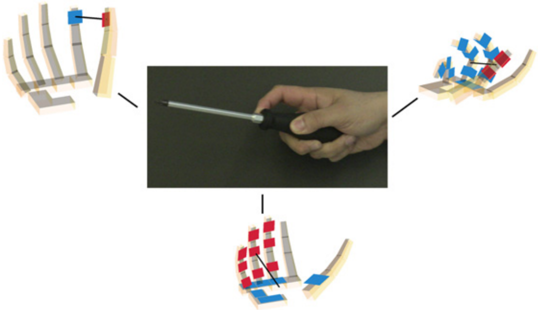

Fig. 2. A screw-driving grasp may be interpreted in terms of 3 oppositions between hand-parts. Each opposition serves a particular functional role. Action of the thumb against side of the fingers supports the action of fingers against the palm in order to keep the tool gripped firmly. Use of the thumb-tip against the finger-tip enables the tool to be directed appropriately during the task.

This paper makes the following contributions:

- We extend the definition of Opposition Space in [4] so that the full flexibility of the thumb – in opposing finger surface, palm and finger sides – may be recognized as separate components of a demonstrated grasp. Additionally, Opposition Space concepts are redefined so that they can be more readily applied in a demonstration context.
- We propose a new way to look at raw information from a grasp demonstration (configuration, tactile) by quantifying the importance of pair-wise interactions between elementary grasping surfaces of the hand. This intermediate representation integrates both hand configuration and interaction force, highlighting the multiple roles that a single sensor patch may have in a grasp. This representation is better able to discriminate among different kinds of oppositions and serves as a basis from which their presence in a grasp may be inferred.
- Inference of the most likely oppositions is done automatically without the use of heuristics. The method is not tied to any functional category of grasps. Experiments with human demonstrations show that the proposed method allows for recognition of a wide range of human grasp intentions with a recognition rate of 87%.

2. Related work

Different approaches are adopted in the literature when seeking to represent information from demonstrated human prehensile posture. Here we consider three approaches that are commonly encountered: joint angles and joint synergies, discrete classification of hand function through grasp taxonomies, virtual fingers in opposition.

Studies in human motion [6,16] have shown that there exists significant correlation among finger joint movements in prehensile postures for everyday tasks. Grasp configurations may therefore be represented by a low-dimensional subspace of a few principal components, known as hand synergies. The small number of dimensions makes it feasible to search for grasp configurations using metrics for overall grasp stability as presented in [17]. However, synergy representations face problems in task related scenarios where specific hand configurations appropriate for the functional requirements of the task are required, such as the examples shown in Fig. 1.

Alternatively, one may start with a set of grasps representing the functional categories of interest. Functional taxonomies such