Relative information content of gestural features of non-verbal communication related to object-transfer interactions

Ansgar R. Koene (a.r.koene@bham.ac.uk), Juliane J. Honisch (j.j.honisch@bham.ac.uk), Satoshi Endo (s.endo@bham.ac.uk) and Alan M. Wing (a.m.wing@bham.ac.uk)

Department of Psychology, Behavioural Brain Sciences Centre, University of Birmingham

Birmingham, B15 2TT UK

Abstract

In order to implement reliable, safe and smooth human-robot object handover it will be necessary for service robots to identify non-verbal communication gestures in real-time. This study presents an analysis of the relative information content in the gestural features that together constitute a communication gesture. Based on this information theoretic analysis we propose that the computational complexity of gesture classification, for object handover, can be greatly reduced by applying attention filters focused on static hand shape and orientation.

Keywords: gestures; Information Gain; object handover; classification;.non-verbal communication

Introduction

The development of service robots that assist humans in their homes and workplace will require fluent communication between humans and robots. Gestures, in particular, play an important part in human-human interactions. For example, people use hand gestures to convey messages when handing over objects (McNeill, 1992). Gaze direction of a person also provides a useful cue for predicting forthcoming action (Kirchner, Alempijevic, & Dissanayake, 2011) and indicating whether a person is interested in a given activity (Argyle & Cook, 1976). Thus, in order to facilitate smooth, comfortable Human-Robot interaction (HRI), it will be necessary for robots to recognize, and produce, these various gestures (e.g. Riek et al, 2010).

As part of the CogLaboration[1] project, which aims to develop a robotics architecture for fluent HRI, we are studying the gestures that are commonly involved in object handover. Non-verbal communication of requests related to object handover was identified as a key competence requirement for the development of service robots (e.g. Nehaniv et al., 2005, Ou & Grupen, 2010).

Smooth object handover relies on a combination of initial purely communicative gestures indicating a desired action, like “pass it faster” etc. and later adjusting of hand/body posture to indicate desired object orientation and placement during the transfer. Recent models of human-robot object handover have therefore attempted to incorporate non-verbal cues in addition to the psychological aspects of the handover model in order to improve the success and safety of handover (e.g. Grigore et al., 2012) and facilitate its smoothness (Cakmak et al ,2011).

[1] www.coglaboration.eu

In the current study we focused on the purely communicative phase preceding the handover. The communication gestures in this phase can be classified into six main groups (Nehaniv et al., 2005; Strabala et al., 2013, Riek et al. 2010): indicating a desire for the partner to “Halt” (H), “Give the object they are holding” (G), “fetch an indicated, non-held, Object” (O), “Take ‘my’ object” (T), “move Faster” (F) and “move Slower” (S).

Despite great advances in computing power, computer vision algorithms and sensing systems, real-time processing of rich sensory information remains a challenge, especially when the computational hardware is limited by the on-board processing power of a mobile robot. At the same time people are very sensitive to slow responses by their interaction partners with delays leading to frustration and irritation. One of the most effective ways to reduce computational load is to reduce the amount of information that is analyzed, i.e. to apply attention filters that focus only on specific volumes of space and/or sensory features.

In this study we therefore investigated which visual features contribute the most information during the purely communicative phase of object handover, thus qualifying for prioritized processing to achieve real-time behavior. We recorded and scored various features of hand/arm gestures that were produced during non-verbal requests related to object handover. The resulting data was analyzed to compute the relative Information Gain and conditional gesture probabilities associated with the various gestural features. Our primary finding is that close to 80% of the information provided by the participant’s gestures can be found by focusing on the static shape and orientation of the hand(s). This suggests that many computational resources can be saved by discounting regions that are not near the hands, and that movement dynamics need not be considered for the purposes of identifying desired behavioral responses.

Method

Participants

Twelve healthy participants, six males and six females, volunteered for this study (age range 21 -37 years). The participants were: two Greek, two Italian, three British, one Dutch, two Malaysian and two Japanese. All participants with one exception were right-handed. Participants were paid and gave informed consent according to institutional guidelines, including an additional video consent form (Ethics Committee of the University of Birmingham).

Apparatus

One video camera and a 12-camera motion tracking system (Qualisys, Sweden) recorded the kinematics of the hand and arm movements of the participants. Each task instruction was presented, one per trial, on a computer monitor that was placed on the left hand side next to the participant.

Procedure

Participants were told to produce hand/arm gestures to non-verbally communicate a desired object handover related action to the experimenter (co-author JH). The six gesture instructions were: “Halt” (H), “Give ‘your’ object” (G), “give Other object” (O), “Take ‘my’ object” (T), “move Faster” (F) and “move Slower” (S). In order to make the task more realistic a real object (small black cup) was used for the handovers. Depending on the task the object was either held by the experimenter (action (H), (G), (O), (F), (S)) or by the participant (T). Each of the six actions was repeated four times. The order of these actions was pseudo-randomized with avoidance of immediate repetition of the same action. The whole study lasted 30 minutes.

Design and Analysis

Each of the 288 (12 participants x 6 actions x 4 repetitions) recorded gestures was scored for each of the 18 gestural features indicated in figure 1. These features were chosen based on a combination of ease with which they might be identified by a robot and their intuitive appropriateness for communicating the six tasks. All scoring was done by one of the authors (AK) who had no prior knowledge of the action instruction of the particular trials. Scoring of the video sequences produced a 288x20 table indicating the presence/absence of each of the 18 features in each of the 288 videos, with two extra columns to record the subject number and the action that was being non-verbally communicated. The latter was identified by experimenter JH after scoring was complete. In order to identify the relative informativeness of each gestural feature we computed the relative Information Gains (IG) for gesture classification.

Information Gain $\boldsymbol{IG(X, a)}$ is an Information Theoretic measure of the reduction in Information Entropy $\boldsymbol{H(X)}$ of a variable $\boldsymbol{X}$ due to knowing the state of variable $\boldsymbol{A}$ (MacKay D.J.C., 2003)

$$IG(X, a) = H(X) - H(X|A)$$

where Information Entropy, in bits, is

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2(P(x_i))$$

and conditional Information Entropy $H(X|A)$ is

$$\boldsymbol{H(X|A) = \sum_j P(A = a_j) H(X|A = a_j)}$$

$$H(X|A = a_j) = -\sum_{k=1}^{m} p(X = x_k|A = a_j) \log_2 p(X = x_k|A = a_j)$$

IG is a popular measure in data mining for identifying the most efficient order in which to process data to reach a classification decision (Bramer, 2007). $IG(X|A)$ measures the reduction in the average number of yes/no type questions that would be required to identify the correct classification of X assuming we can start with knowing A.

Probability of correct classification. While IG is convenient for summarizing the increase in predictability of variable X, given knowledge of variable A, $IG(X|A)$ alone does not provide direct insight into the probabilities with which we can expect to correctly classify any of the action communication gestures. To gain more insight into expected classification performance we therefore also looked at the conditional probabilities $p(X = x_i|A = a_j)$.

Results

Relative IG for single gestural features

Figure 1 depicts the Information Gain of knowing any one of the single gestural features relative to the Information Gain of knowing all 18 features, $IG(X|single)/IG(X|all)$.

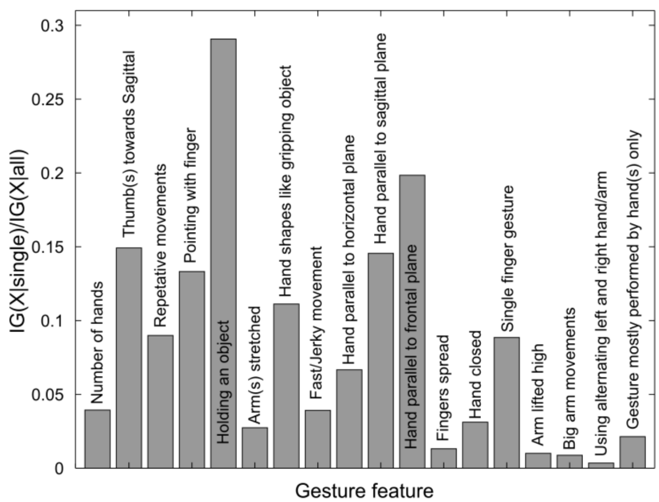

Figure 1: Relative IG for single gestural features.

Based on this analysis, the six most informative features are (1) holding an object, (2) hand parallel to frontal plane, (3) thumb(s) towards mid sagittal plane, (4) hand parallel to sagittal plane, (5) pointing with finger and (6) hand shaped like gripping an object. Of these six features all except “holding an object” are related to hand shape/orientation and none involves movement dynamics. Even “holding an object” is likely to be identifiable if visual processing is focused on hand shape. The most information rich dynamic gesture property is “repetitive movement”, which appears as the 7th most informative feature. Even the most informative feature however had an Information Gain of less than 30% relative to knowing all features. We therefore analyzed the IG for knowing combinations of static hand features.

Relative IG for static hand features

Figure 2 depicts the relative Information Gain of knowing a combination of static hand shape and/or orientation.

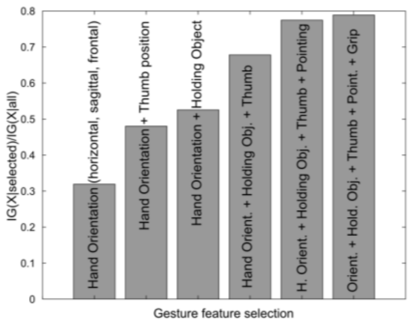

Figure 2: Relative IG for selected combinations of gesture features related to hand orientation and/or shape.

We note that while knowledge of "hand orientation" alone produces just over 30% of the *IG* over all features, adding only additional information concerning the relative position of the thumb (i.e. disambiguating between 180deg torsional rotations of the hand) brings the relative *IG* up to almost 50%. A slightly better performance is achieved by adding information about an object in the participant's hand, while adding both thumb position and object holding raises the relative *IG* to almost 70%. An additional identification of "pointing" behavior raises the relative *IG* to almost 80%, while information about grip like hand shaping has very little additional impact on gesture classification.

Conditional probabilities of correct classification

Figures 3-6 show the conditional probabilities for the six task conditions (Take, Give, Other object, Faster, Slower, Halt) given knowledge about gestural features. The conditional probabilities $p(X = x_i|A = a_j)$ were computed directly from the score table as the ratio of trials with feature a_j that belong to task condition x_i over the total number of trials with that gestural feature.

$$p(X = x_i|A = a_j) = \frac{\#trials(X = x_i, A = a_j)}{\#trial(A = a_j)}$$

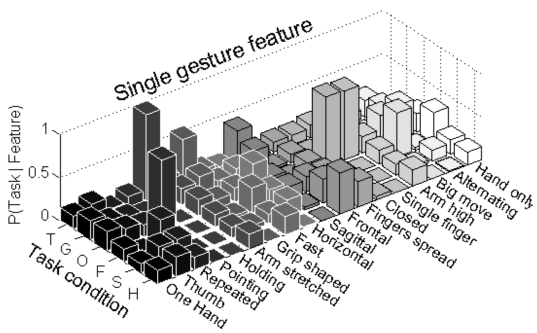

Figure 3: Conditional task probabilities given knowledge of a single gesture feature.

The results in figure 3 clearly reveal that the gestural features "holding an object", "pointing", "closed hand" and "single finger" are highly indicative of specific task conditions at p(X|A)=1, 0.81, 0.81 and 0.79, respectively. The latter two features do not have high *IG* however due to their low probability of occurrence ($p(single\ finger) = 0.1; p(closed\ hand) = 0.06$).

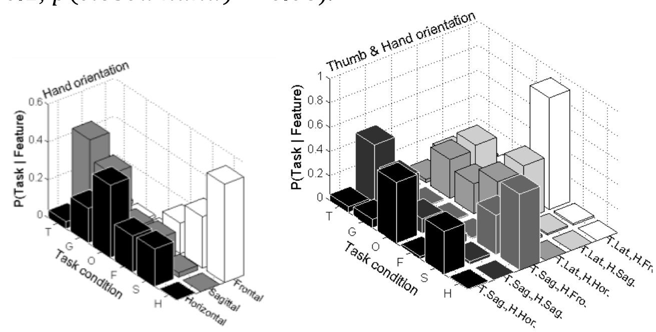

Figure 4: Conditional task probabilities given knowledge of (A) only hand, (B) thumb and hand orientation.

Figure 4A focuses in on the hand orientation features, showing how fronto-parallel orientation is clearly indicative of gestures related to movement speed (Faster, Slower, Halt) while sideways tilted hand(s) (sagittal orientation) are generally related to requests for initiation of interaction (Take, Give, pass the Other). Horizontal hand gestures however appear to fall in both categories.

The addition of knowledge about the relative location of the thumb (figure 4B) disambiguates between palm up, down inwards or outwards, separating the fronto-parallel hand gestures into a "Faster" group and the "Slower or Halt" group. Figure 5 finally, when compared to figure 4B, shows how the high *IG* feature of "holding an object" greatly improves the probability of identifying "Take mine" but makes no obvious contribution towards disambiguating any of the other gestures.

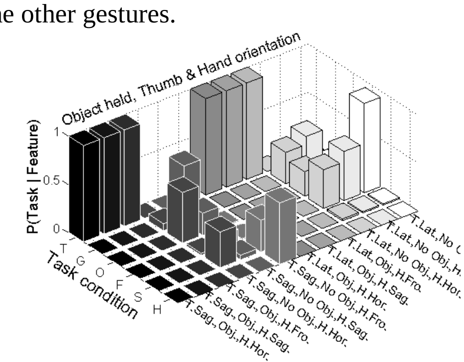

Figure 5: Conditional task probabilities given knowledge of thumb and hand orientation and presence of held objects.

Discussion

In this paper we set out to evaluate the relative information gains that can be had from observing various gestural features. The ultimate purpose is to identify those features that should be prioritized for reliable gesture recognition, under the temporal and computational limitations of real-time processing by service robots.

The gesture tasks in this study focused on non-verbal communication of behavior requests related to object handover. Based on the relative frequency with which our participants produced each of 18 gestural features to signal six handover related request, we computed the information theoretic measure of (relative) Information Gain (IG) for each feature, and various combinations thereof. The *IG* analysis revealed that close to 80% of the information that would be gained by knowing the full 18 features could be achieved purely from static hand and thumb posture information in combination with detecting if the participant is holding an object. This result was further supported by the strongly peaked conditional probability distributions for identifying the task condition, given knowledge of this subset of gestural features (figure 5), indicative of (relatively) unambiguous classification.

Focusing of computational resources on the small spatial regions of the hands, while classifying only static hand shape and orientations, holds the promise of greatly reducing the computational load and required time windows of information gathering. By not having to track dynamic movement aspects it is possible to work at lower frame rates and avoid the use of time derivatives (velocity, acceleration, jerk) of the information time series, which are increasingly prone to noise.

It should be noted however that there are a number of caveats that need to be addressed in follow up studies. Firstly, this study has focused only on the issue of determining which visual cues to prioritize for real-time gesture recognition. This does not address all possible techniques for improving rapid non-verbal communication recognition. Gesture classification will obviously be greatly improved by including various types of prior knowledge, e.g. knowledge that gestures for "halt" or "slower" are much less likely to occur when the handover action has not (yet) started. Secondly, in the current experiments, identifying if the partner is "holding an object" is very highly predictive of a "take 'my' object" gesture because participants were only holding objects for that specific task condition. By the nature of the "take object" condition this will probably also be true in most realistic settings, but not always. For natural settings, the predictive power of "holding an object" should therefore be considered as overestimated in our current study.

Another issue that will require further study concerns variability in inter subject behavior. Even among our 12 participants some task conditions produced a wide variability in the gestures. One possible source of this variability may be cultural differences, since the participants in our study originally came from both northern and southern Europe as well as Asia. The general importance of cultural differences for non-verbal communication is well recognized (Morris, et. al., 1979). Gesture recognition performance might therefore be greatly improved through region specific tuning. In recognition of this possibility we are now engaged in an internet based study[2] to have people from across the world rate how they would interpret (re-enactments of) the gestures we observed in our participants.

Finally, as part of the Behavior Informatics project, aimed at increasing accessibility to behavior related data, we are making our labeled and scored database of object transfer gestures (the re-enactments used in our on-line study) available for download[3].

[2]http://symonlabexperiments.site90.net/GesturesIntroPage.html

Acknowledgments

This work was supported by the CogLaboration project within the 7th Framework Programme of the European Union, - Cognitive Systems and Robotics, contract Number FP7-287888, see also http://www.coglaboration.eu/

References

Argyle, M. & Cook, M. (1976). *Gaze and mutual gaze.* Cambridge, UK: Cambridge University Press.

Bramer, M. (Eds.) (2007), *Principles of Data Mining*. London, UK: Springer.

Cakmak, M., Srinivasa, S.S., Lee, M.K., Kiesler, S., & Forlizzi, J. (2011). Using spatial and temporal contrast for fluent robot-human handovers. *Proc. 6th int. conf. on Human-robot interaction* (pp. 489-496).

Grigore, E.C., Eder, K., Lenz, A., Skachek, S., Pipe, A.G., & Melhuish, C. (2011). Towards safe human-robot interaction. *12th Conference Towards Autonomous Robotic Systems* (pp. 323–335). LNCS 6856, Springer.

Kirchner, N., Alempijevic, A., & Dissanayake, C. (2011). Nonverbal robot-group interaction using an imitated gaze cue. *Proc. 6th int. conf. on Human-robot interaction (HRI '11)* (pp. 497-504). New York: ACM.

MacKay, D.J.C. (2007), *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press.

McNeill, I.D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Morris D., Collett P., Marsh P. and O'SHaughnessy M. (1997). *Gestures: their origins and distribution*. New York: Stein & Day.

Nehaniv, C., Dautenhahn, K., Kubacki, J., Haegele, M., Parlitz, C., & Alami, R. (2005). A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction, *Proc. IEEE Int. Workshop on Robot and Human Interactive Communication* (pp. 371-377).

Ou, S., & Grupen, R. A. (2010). From manipulation to communicative gesture. *Proc. of the 5th ACM/IEEE Int. Conf. on Human Robot Interaction (HRI 2010)*.

Riek, L.D., Rabinowitch, T.C., P., Bremner, A., Pipe, Fraser, M., & Robinson, P. (2010). Cooperative gestures: effective signaling for humanoid robots. *Proc. of ACM Int. Conf. on Human-Robot Interaction.*

Strabala, K., Lee, M.K., Dragan, A., Forlizzi, J., Srinivasa, S., Cakmak, M., & Micelli, V. (2013). Towards Seamless Human-Robot Handovers. *Journal of Human-Robot Interaction, 1(1),* 1-23.

[3]http://behaviorinformatics.bham.ac.uk/index.php/CogLab,_Object_Transfer_Gestures