Human Gesture Recognition using Dual Quaternion Algebra

Ana Christina Almada Campos and Bruno Vilhena Adorno

***Abstract*—This work proposes an approach for interpretation of human pointing gestures using geometric primitives represented by dual quaternions. Objects are modeled with planes and a line represents the gesture. The crossing point between line and plane determines the indicated object. Experimental results show a mean success rate of 77,78% (standard deviation of 16,41%) in the system's interpretation.**

I. INTRODUCTION

Human-robot interaction has been of great interest in recent years, with the increasing application of robots in human environments. Efficient communication between agents is a fundamental aspect of a good interaction, especially in collaboration contexts. Human-human interaction provides insights of how a robot can understand human actions, even implicit cues, and communicate efficiently with a human partner [1], [2]. Common ways of communication include both explicit (when information is passed intentionally) and implicit ones (when there is no intention to communicate). An agent can implicitly anticipate commands from partners only by observing what is their attention focus or which movement is being executed. Meanwhile, explicit commands can be given by voice or gestures.

A robot in a collaborative or assistive application needs to interpret human commands to execute the required action. We propose an approach for interpretation of human gesture commands, using dual quaternion algebra and geometric primitives.

II. METHODS

A. Pointing line

We used a Microsoft Kinect to obtain information about human joints. The application tracks fifteen of them, including right and left hips, elbows, and hands. The distances between the hands and hips were used to determine if the user was in a position of standing or pointing. If pointing, the greater hip-hand distance indicates which arm is being used to the gesture. The joint tracking application provides information about position and orientation of each joint frame, which were then expressed using unit dual quaternions. Elbow and hand poses are given by the unit dual quaternions $\underline{\boldsymbol{x}}_{\text{elbow}}$ and $\underline{\boldsymbol{x}}_{\text{hand}}$, respectively.

Ana Christina A. Campos is with the Graduate Program in Electrical Engineering - Universidade Federal de Minas Gerais (UFMG) - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte-MG, Brazil. Email: anachristinaac@gmail.com.

Bruno V. Adorno is with the Department of Electrical Engineering - Universidade Federal de Minas Gerais (UFMG) - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte-MG, Brazil. Email: adorno@ufmg.br.

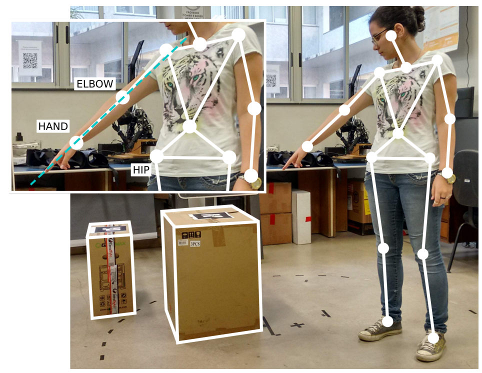

Fig. 1. Illustration of a pointing gesture. The right hip-hand distance is larger than a predefined threshold and than the left hip-hand distance, indicating that the right arm is being used for the gesture. Elbow and hand points are used to create a line (*dashed green*) representing the gesture.

We can extract the position quaternion $\boldsymbol{p}$ from the unit dual quaternion $\underline{\boldsymbol{x}} \in \text{Spin}(3) \ltimes \mathbb{R}^3$ by doing $\boldsymbol{p} = 2\,\mathcal{D}\left(\underline{\boldsymbol{x}}\right)\mathcal{P}(\underline{\boldsymbol{x}})^*$, where $\mathcal{P}(\underline{\boldsymbol{x}})$ and $\mathcal{D}\left(\underline{\boldsymbol{x}}\right)$ operators return the primary and dual parts of $\underline{\boldsymbol{x}}$, respectively, and $\mathcal{P}(\underline{\boldsymbol{x}})^*$ indicates the conjugate of $\mathcal{P}(\underline{\boldsymbol{x}})$, which is analogous to the conjugate of a complex number [3]. We define a unit pure dual quaternion Plücker line $\underline{\boldsymbol{l}}$ to represent the gesture using elbow and hand positions. The unit pure quaternion $\boldsymbol{l}$, given by $\boldsymbol{l} = (\boldsymbol{p}_{\text{hand}} - \boldsymbol{p}_{\text{elbow}})/\left\|(\boldsymbol{p}_{\text{hand}} - \boldsymbol{p}_{\text{elbow}})\right\|$, corresponds to the line direction and the gesture line is given by $\underline{\boldsymbol{l}} = \boldsymbol{l} + \varepsilon(\boldsymbol{p}_{\underline{\boldsymbol{l}}} \times \boldsymbol{l})$, where $\boldsymbol{p}_{\underline{\boldsymbol{l}}} \in \{\boldsymbol{p}_{\text{hand}}, \boldsymbol{p}_{\text{elbow}}\}$. Fig. 1 illustrates how to obtain the pointing line from human joints.

B. Objects models

We represented the objects as boxes composed of six planes and assumed that planes are parallel or perpendicular to the axes of the inertial frame. Let $\boldsymbol{n}_i$ be a unit norm pure quaternion normal to the plane $\underline{\boldsymbol{\pi}}_i$, with $i \in \{1, \ldots, 6\}$, $\boldsymbol{q}_i$ an arbitrary point on it and $\langle \boldsymbol{q}_i, \boldsymbol{n}_i \rangle \triangleq -(1/2)\left(\boldsymbol{q}_i\boldsymbol{n}_i + \boldsymbol{n}_i\boldsymbol{q}_i\right)$ the dot product between them [3]. Thus, each plane $\underline{\boldsymbol{\pi}}_i = \boldsymbol{n}_i + \varepsilon\langle \boldsymbol{q}_i, \boldsymbol{n}_i \rangle$ is defined with its normal pointing inside the object.

C. Gesture interpretation

We check if a point is on the plane $\underline{\boldsymbol{\pi}}$ by calculating the projection of the point onto the normal $\boldsymbol{n}$ and comparing it with the dual part of the plane $d_{\underline{\boldsymbol{\pi}}} = \langle \boldsymbol{q}, \boldsymbol{n} \rangle$. We use that to find the crossing points $\boldsymbol{p}_c$ of the gesture line and the object's planes. Starting from the point on the line $\boldsymbol{p}_{\underline{\boldsymbol{l}}}$, we have

$$\boldsymbol{p}_c = \boldsymbol{p}_{\underline{\boldsymbol{l}}} + D_c \boldsymbol{l}, \tag{1}$$

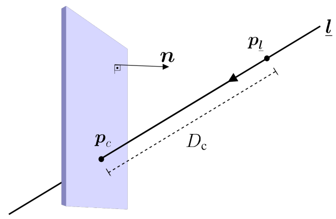

Fig. 2. The crossing point between line and plane is used to determine if the object is a possible indicated one. The crossing point $\boldsymbol{p}_c$ and the distance on line D_c are obtained using (1) and (3).

where D_c is the distance from $\boldsymbol{p}_{\underline{l}}$ to $\boldsymbol{p}_c$ on the line. Fig. 2 illustrates the procedure. The crossing point needs to satisfy

$$\langle \boldsymbol{p}_{\underline{l}} + D_c \boldsymbol{l}, \boldsymbol{n} \rangle = \langle \boldsymbol{p}_{\underline{l}}, \boldsymbol{n} \rangle + D_c \langle \boldsymbol{l}, \boldsymbol{n} \rangle = d_{\underline{\boldsymbol{\pi}}}. \tag{2}$$

Rearranging (2), the distance D_c can be found as

$$D_c = \frac{d_{\underline{\boldsymbol{\pi}}} - \langle \boldsymbol{p}_{\underline{l}}, \boldsymbol{n} \rangle}{\langle \boldsymbol{l}, \boldsymbol{n} \rangle}. \tag{3}$$

Thus, the distance of this crossing point to all the six planes is calculated using the function $d(\boldsymbol{p}_c, \underline{\boldsymbol{\pi}}_i) : \mathbb{H}_p \times \mathcal{H} \to \mathbb{R}$, such that $d(\boldsymbol{p}_c, \underline{\boldsymbol{\pi}}_i) = \langle \boldsymbol{p}_c, \boldsymbol{n}_i \rangle - d_{\underline{\boldsymbol{\pi}}_i}$, with $i \in \{1, \ldots, 6\}$. Distances are measured from the planes and, if they are all non-negative, the crossing point is inside the object, which is then added as a possible indicated one. If this procedure returns more than one object, the ambiguity is solved by verifying the distance D_c. Since the shortest distance corresponds to the object that it is closer to the point $\boldsymbol{p}_{\underline{l}}$ located on the person, that object is considered the person's indication, as the further ones are occluded by it.

III. Experimental Results and Discussion

We tested the proposed method in an experiment with objects represented by boxes of known dimensions and locations, as illustrated in Fig. 1. The planes were defined considering dimensions 20% larger than the real ones to guarantee that the real object was inside the virtual model. We asked volunteers to execute a sequence of ten pointing gestures that could indicate one of four objects or none of them. The sequence was chosen by the subjects and they informed it to the experimenter after the execution. The experimenter was responsible to document the system's answers and the experiment was filmed with the knowledge and agreement of the participants. Eleven subjects participated, but two of them were excluded after the video analysis. These subjects did gestures to indicate the objects without pointing to them. Instead, they only raised their arms over the objects. Also, one of the subjects provided a reference sequence that was not the one executed. Therefore, this reference was corrected to be in accordance with the video.

The results are shown in Fig. 3. The mean success rate was $77,78\%$, with standard deviation of $16,41\%$. Considering all gestures, the false negative rate, i.e., when the system did not detect an object and the subject was pointing at one, was $32,20\%$. A case of false positive, i.e., when the system detected an object and the subject was not pointing to any of them, did not occur. Only once the system indicated an object while the subject was pointing to a different one, and this case

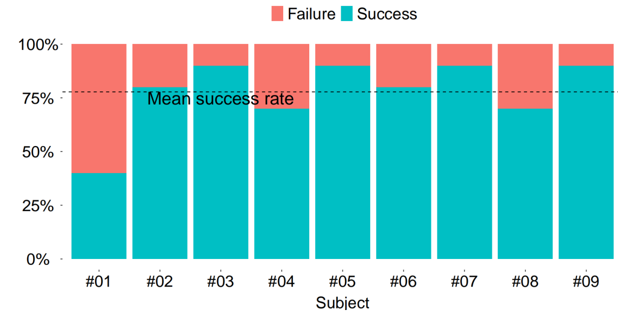

Fig. 3. Experimental results. Columns indicate the success and failure rates of each subject and the dashed line indicates the mean success rate.

involved objects 3 and 4, which were intentionally positioned to represent an ambiguity: depending on the point of view of the subject, object 3 was occluded by object 4. In two other gestures, the system oscillated between considering object 3 or object 4 and in these cases, if the correct answer was one of them, it was counted as a success.

We observed that most failures were related to object 3. There was a blind spot that made it difficult to point at object 3 and be recognized by the system. The Kinect presented an offset in measurements that we compensated by a transformation defined experimentally. We believe that the blind spot was caused by the uncertainty of the measurements. Other errors occurred because of occlusion of the pointing arm. We attribute the remaining failures to the fact that the system indicates no object even if the pointing line is very close to one of them. However, in human-human communication, the interpretation is not so rigorous and this can be relaxed.

IV. Conclusions

This work proposed an approach for human gesture interpretation, tested in an experiment to detect the object that subjects pointed to. The results show that simple solutions, such as the inclusion of more Kinects, could improve system's performance. In applications where the robot interacts with the same group of people, they get familiar with the system and the performance can also improve. The same goes for a personal assistive robot application. Furthermore, while interpreting gestures, humans can use other types of information, such as eye gaze. The same approach presented here is already being tested in the detection of the human attention focus based on eye gaze. Therefore, future works include system's improvement by crossing different information, using the same method presented here.

References

[1] N. Gildert, A. G. Millard, A. Pomfret, and J. Timmis, "The Need for Combining Implicit and Explicit Communication in Cooperative Robotic Systems," *Frontiers in Robotics and AI*, vol. 5, no. JUN, pp. 1–6, jun 2018.

[2] J. Shah and C. Breazeal, "An Empirical Analysis of Team Coordination Behaviors and Action Planning With Application to Human-Robot Teaming," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 2, pp. 234–245, apr 2010.

[3] B. V. Adorno, "Robot Kinematic Modeling and Control Based on Dual Quaternion Algebra – Part I: Fundamentals," 2017. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01478225v1