

# Operator Awareness in Human-robot Collaboration through Wearable Vibrotactile Feedback

Andrea Casalino<sup>1\*</sup>, Costanza Messeri<sup>1\*</sup>, Maria Pozzi<sup>2\*</sup>,  
Andrea Maria Zanchettin<sup>1</sup>, Paolo Rocco<sup>1</sup>, and Domenico Prattichizzo<sup>2</sup>

**Abstract**—In industrial scenarios requiring human-robot collaboration, the understanding between the human operator and his/her robot co-worker is paramount. On one side the robot has to detect human intentions, and on the other side the human needs to be aware of what is happening during the collaborative task. In this paper, we address the first issue by predicting human behaviour through a new recursive Bayesian classifier exploiting head and hand tracking data. Human awareness is tackled by endowing the human with a vibrotactile ring that sends acknowledgements to the user during critical phases of the collaborative task. The proposed solution has been assessed in a human-robot collaboration scenario and we found that adding haptic feedback is particularly helpful to improve the performance when the human-robot cooperation task is performed by non-skilled subjects. We believe that predicting operator's intention and equipping him/her with wearable interfaces able to give information about the prediction reliability, are essential features to improve performance in human-robot collaboration in industrial environments.

**Index Terms**—Human-Centered Robotics, Haptics and Haptic Interfaces, Cognitive Human-Robot Interaction

## I. INTRODUCTION

HUMAN-ROBOT collaboration (HRC) has received increasing attention over the last years, [1]. The possibility for robotic manipulators to share the workspace with humans is becoming an important asset, especially in small and medium sized enterprises (SMEs), and the corresponding applications are rapidly contaminating other fields such as healthcare and domestic robotics.

For a long time, safety has been the main focus in HRC, [2]. More recently, researchers have started working on methods to allow a fluent interaction between humans and robots. The possibility for a properly instrumented robotic device to understand and somehow predict humans' intentions is now considered as important as safety. In this context, many results have been reported showing the increasing capability of robots to semantically interpret their human fellows. For example, in

[3], a method based on conditional random fields (CRFs) is used by the robot to anticipate its assistance. In Luo et al. [4] Gaussian Mixture Models (GMMs) are used to predict human reaching targets. A similar result has been obtained in [5] by means of Bayesian methods. In [6] human intention is inferred by combining expectation-maximisation (EM) algorithms and an online model learning strategy.

However, allowing robots to understand humans is by far not enough to establish a fluent and effective collaboration. In fact, recent papers have motivated the need for mutual understanding, where the human has to be aware about the inference capabilities of the robot, [7], [8]. Technological advancements have introduced a plethora of new methodologies to increase the awareness of the operator during the collaboration with a robot. For example, in [9] an augmented reality (AR) has been introduced to support the human in collaborative assembly operations. Signal lights and their optimal positioning have been addressed in [10] to inform the human operator about the status of the robot. Verbal feedback, i.e. the most natural (for the human) interaction modality, has been addressed in [11] showing its capability to improve the performance in HRC. Acoustic and visual feedback signals are employed in HRC also to let the human know if his/her commands to the robot were successfully recognized [12].

In this paper, we propose a HRC framework where the human intention is estimated by means of a Bayesian recursive classifier, and the human awareness is improved by using haptic feedback.

The sense of touch is not only the most robust and distributed of human senses, but it is also proximal, bidirectional, and private [13]. These features make the haptic channel particularly suitable to convey information to workers in industrial environments, where visual and auditory modalities might be busy to effectively accomplish a task (e.g., the human must use the view to find and pick objects), or impaired due to personal protective equipment (PPE) (e.g., the worker could be wearing headphones).

Recently, the new generation of wearable haptic interfaces [14] has widened the range of the possible applications of haptics in many areas, including robotics. Wearable haptics combines the advantages of haptic feedback with those of wearability, a key feature that “enables novel forms of communication, cooperation, and integration between humans and machines” [15]. Devices that can be worn on different parts of human body (hand-exoskeletons, armbands, rings, thimbles) and that can transmit different types of haptic feedback (cutaneous and kinaesthetic force feedback, vibrations) have been

Manuscript received: February, 24, 2018; Revised June, 9, 2018; Accepted July, 14, 2018.

This paper was recommended for publication by Editor Yasuyoshi Yokokohji upon evaluation of the Associate Editor and Reviewers' comments.

<sup>1</sup>A. Casalino, C. Messeri, A.M. Zanchettin and P. Rocco are with Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Milano, Italy. {andrea.casalino, costanza.messeri, andreamaria.zanchettin, paolo.rocco}@polimi.it

<sup>2</sup>M. Pozzi and D. Prattichizzo are with the Department of Information Engineering and Mathematics, University of Siena, Italy and the Advanced Robotics Department, Istituto Italiano di Tecnologia, Genova, Italy. pozzi@diism.unisi.it, prattichizzo@ing.unisi.it

\*equally contributed to this work.

Digital Object Identifier (DOI): see top of this page.

tested in various contexts. Haptic thimbles and bracelets were successfully applied in human-robot cooperation scenarios such as teleoperation [16] and human-robot teams navigation [17]. Haptic rings were found to be particularly suitable in augmented reality applications, where humans have to interact not only with a virtual environment, but also with real objects [18], and in conjunction with a wearable extra robotic finger, both to control it and to perceive tactile feedback from it [19], [20]. The most widespread tactile displays are those that generate vibrotactile sensations [21]. Several commercial products, including mobile phones, gaming controllers, car cockpits, and surgical instruments, employ vibrations to transmit tactile sensations to the users, or just to alert them that a certain event occurred [13].

In this paper, we tackle the problem of human-robot collaboration with a bidirectional information exchange. On the one hand, an intention estimation algorithm tracks the human, and allows the robot to understand in advance the forthcoming action of the operator. On the other, a wearable device is used to improve the operator awareness during the human-robot collaborative assembly task through vibrotactile feedback. The proposed inference method significantly extends the one in [5] considering the motion of both human hands and also the head orientation. Moreover, to the best of our knowledge, this is the first time that a haptic interface is used to augment the collaboration abilities between humans and robots.

The rest of the paper is organized as follows. Section II explains the intention prediction algorithm, Section III shows how the method can be applied to a real HRC scenario, and Section IV analyses the experimental results.

## II. METHODS

### A. Generalities

To achieve an efficient human-robot collaboration, it is essential for the robot to infer human's intention and then decide the best action to take. The difficulty of inferring human intentions is proportional to the level of abstraction associated to the possible activities that the operator can undertake. For instance, when monitoring an operator in a robotic cell, we can just infer which tool the human intends to reach, or solve the more difficult problem of inferring what part he/she is going to assemble within a set of possible ones (activity labelling). The second kind of problems can be solved by using the approach proposed in [22] and [3], adopting CRFs, however relying on a complex training phase. Moreover, the learnt model will be valid only for the specific environment monitored in the training phase.

By reducing the level of abstraction, we can make inference in more general contexts, relying on a reduced a priori knowledge. In this work, we focus on inferring the next goal that will be reached by the operator's hand, among a set of possible known ones. One common approach is to monitor the whole operator's arm trajectory [4]. Indeed, the observed trajectory of the hand, as well as other skeletal points of interest, are treated as features, which are exploited to classify the trajectory using a learnt GMM. After classification, it is possible to forecast the trajectory on a small future horizon

time. The predicted motion can be then exploited to make inference about the target location of the human.

Zanchettin and Rocco, in [5], considered only the operator's hand trajectory and solved the problem by using a recursive Bayesian classifier. In this work, we develop a similar method, however considering as observations both, the position of the operator's hands and the orientation of the operator's head, which is an estimate of his/her gaze. All the aforementioned quantities can be acquired through a motion sensing input device.

Monitoring the operator's gaze is not something new in human-robot interaction (HRI). It is used, for example, to let social robots perform human-like motions. In [23], authors focused on producing coordinated head-arm motions for a humanoid robot with a two degrees-of-freedom head.

Even for human-human interactions, the gaze constitutes an important measure, since it is closely tied to what people are thinking and doing [24]. Motivated by this fact, we adopted the estimate of the gaze (face orientation) to help in the process of inference. In a similar way, Bednarik *et al.*, in [25], developed a classifier that is able to detect when a person observes an intended goal location. The classifier is adopted to avoid the so called Midas touch problem, *i.e.*, considering as intended every goal that at a certain time is in the operator's field of view. We think that the same problem can be mitigated by considering also other observations rather than only the gaze, as for example one hand trajectory. This approach was followed in [26], where a hidden Markov model (HMM) was adopted. The HMM allows to make inference about a sliding temporal window of observations.

### B. Probability distributions

In this work, the probability distribution of a set of possible goals that an operator can reach, is computed in a recursive way, as done in [5]. We assume the set of possible goals  $G = \{Goal_1, \dots, Goal_n\}$  is known as well as their locations in space. Regarding the measures retrieved from a depth camera, we will denote with  $\mathbf{p}_k^R$  and  $\mathbf{p}_k^L$  the position of the right and left wrist of the operator at step  $k$ , while we will use  $z_k$  to denote the estimated orientation of the head.  $\mathbf{h}_k$  will denote the estimated position of the skeletal point associated to the head. The sensing device computes  $z$  as the normal of a plane which interpolates some detected facial points (Fig. 1a). Not only the estimate of  $z_k$  can be returned, but a flag  $f_k$  about its validity is returned too. When the operator looks towards a direction that is very different from the normal of the camera frame  $z_{Kin}$  (Fig. 1b), the retrieved measure is a number with no sense. In such cases,  $f_k$  will be returned as false. Let us define a vector  $\theta_k$  containing the complete set of measurements (Fig. 1b):

$$\theta_k = (\mathbf{p}_k^R \quad \mathbf{p}_k^L \quad \mathbf{h}_k \quad z_k \quad f_k)^T.$$

According to the past acquired samples for  $\mathbf{p}^R$  and  $\mathbf{p}^L$ , it is possible to estimate the direction of the velocity of both hands at step  $k$ , which is expressed by the unit vectors  $\mathbf{t}^R$  and  $\mathbf{t}^L$ . Since a task involves the simultaneous motion of both human hands, it is in principle possible to compute two distinct

distributions of probability: one related to the right hand and one to the left hand. Considering one of the aforementioned distributions, the probability of goal  $i$  at step  $k$  will be denoted as  $P_k^i$ . This quantity is interpreted as an a priori knowledge for step  $k+1$ . It is possible to compute the likelihood function of goal  $i$  at step  $k+1$ , *i.e.*,  $\Pi(\theta_{k+1}|Goal_i)$ , by using of the retrieved measures  $\theta_{k+1}$ . This quantity is employed to compute the a posteriori probability at step  $k+1$ , applying the Bayes formula:

$$P_{k+1}^i = \frac{P_k^i \Pi(\theta_{k+1}|Goal_i)}{\sum_{j=1}^n P_k^j \Pi(\theta_{k+1}|Goal_j)}, \quad (1)$$

where  $n$  is the number of possible goals. The key question is how to compute the likelihood  $\Pi(\theta_{k+1}|Goal_i)$ . To this purpose, the approach proposed in [5] computes the angle between  $t^R$  (or  $t^L$ ) and a predicted direction, computed by considering a third degree polynomial path connecting the previous position of the operator's hand (right or left) with  $Goal_i$ . In this way the measure about one single operator's hand is considered. To tackle the same problem, here we propose to compute the likelihood by using a Gaussian Mixture distribution, which considers as features the quantities contained in vector  $\Phi^i_k$ , defined as follows (see Fig. 1b):

$$\Phi^i_k = (\theta_{W1k}^i \quad \theta_{W2k}^i \quad \theta_{Hk}^i \quad d_k^i)^T.$$

When considering inference for the right hand,  $\theta_{W1k}^i$  is taken as the angle between  $t^R_k$  and the vector connecting  $p^R_k$  to the position of  $Goal_i$ , while  $\theta_{W2k}^i$  is a similar angle but considering  $t^L_k$ . On the contrary, when inference is made for left hand,  $\theta_{W1k}^i$  is computed according to  $t^L_k$  and  $\theta_{W2k}^i$  according to  $t^R_k$ . The angle  $\theta_{Hk}^i$  is the one between  $z_k$  and the vector connecting  $h_k$  to the location of  $Goal_i$ , while  $d_k^i$  is the distance between point  $p^R$  and  $Goal_i$ , when considering inference for the right hand while is the same involving  $p^L$  when considering inference for the left one. Angle  $\theta_{W2k}^i$  is included in  $\Phi^i_k$  because a high evidence that left hand is going to  $Goal_i$ , reduces the probability that this goal is intended for the right hand. Clearly we assume that the operator uses a single hand to reach a certain goal<sup>1</sup>. On the contrary, not considering  $\theta_{W2k}^i$  in  $\Phi^i_k$  would lead to update the probability distributions related to the hands in a completely independent way, which is not realistic.

### C. Learning the likelihood function

The distribution of the likelihood function is approximated with a GMM. The GMM probability density function is indicated as  $f(\Phi)$ , and it is basically a weighted sum of multivariate Gaussian densities:

$$f(\Phi) = \sum_{c=1}^{Nc} w_c |2\pi\Sigma_c|^{-0.5} \Upsilon(\Phi|\mu_c, \Sigma_c), \quad (2)$$

with:  $\Upsilon(\Phi|\mu_c, \Sigma_c) = \exp(-0.5(\Phi - \mu_c)^T \Sigma_c^{-1} (\Phi - \mu_c))$  and  $\sum_{c=1}^{Nc} w_c = 1$ .

In Eq. (2),  $Nc$  is the number of clusters of the mixture, while  $\Sigma_c$ ,  $\mu_c$ , and  $w_c$  are the covariance, the mean, and the weight

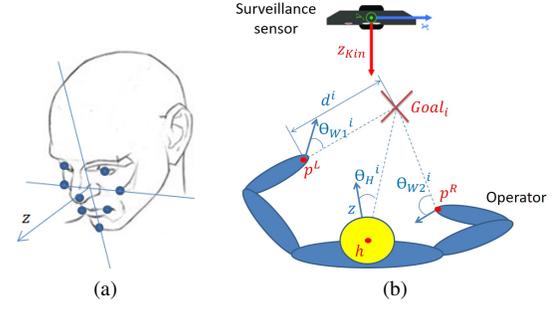


Fig. 1: On the left, estimation of  $z$  is made according to some detected facial points, which are depicted as blue points. On the right all the measured retrieved from the sensor, which are exploited for inference are reported.

of the  $c^{th}$  cluster, respectively. The density  $f$  is exploited to compute the likelihood function:

$$\Pi(\theta_k|Goal_i) = f(\Phi^i(\theta_k, Goal_i))$$

and can be learnt from data in a supervised manner. The parameters to be learnt can be grouped in a set  $\Theta = \langle w_1, \Sigma_1, \mu_1, \dots, w_{Nc}, \Sigma_{Nc}, \mu_{Nc} \rangle$ . We can monitor operators during the execution of certain tasks, collecting some trajectories  $p^R(t)$ ,  $p^L(t)$ ,  $h(t)$ ,  $z(t)$ . Then, by knowing the intended goal locations<sup>2</sup>, it is possible to extract from the collected data a population of samples  $X = \langle \Phi_1, \dots, \Phi_m \rangle$ , which can be used to learn the underlying GMM distribution.

To this purpose we make use of the well known Expectation Maximization (EM) algorithm [27]. It is an iterative algorithm that starts from an initial guess  $\Theta^0$ , computed by using a k-means classifier, and then adjusts values for  $\Theta$  until the convergence to a maximum for  $L$ , which is the log-likelihood of the parameters with respect to sample  $X$  (*i.e.*, a measure of how well the GMM distribution explains the training dataset) defined as:

$$L^k = L(\Theta^k|X) = \frac{1}{m} \sum_{i=1}^m \log\left(\sum_{c=1}^{Nc} w_c^k \Upsilon(\Phi_i|\mu_c^k, \Sigma_c^k)\right).$$

The computation of  $\Theta^{k+1}$  is made so to increase the value of  $L$ , *i.e.*, so to have  $L^{k+1} > L^k$ , till the convergence to a maximum  $L^*$  (see [27]). The number of clusters defining the GMM distribution, *i.e.*, the cardinality of  $\Theta$ , must be decided prior to applying EM. Anyway it is possible to apply the EM more than once, changing the number of clusters every time. At the end, it is possible to compare the maximum value  $L^*$ , achieved with different numbers of clusters to determine the best one among all. Indeed, this was the procedure used in this work.

We conclude this section by pointing out that the learnt model is general, in the sense that it can be applied to any kind of robotic cell (with its own locations for goals). In this way, when changing the layout of a cell, we can still rely on the learnt GMM, without a new training of the model. On the other hand, a specific GMM learnt for a newer layout could achieve higher performance.

<sup>1</sup>But moving simultaneously both hands to reach different locations.

<sup>2</sup>In this sense the approach is supervised.

#### D. Making inference with GMM

Once the GMM describing the likelihood function  $\Pi(Goal_i|\theta)$  has been learnt, it is possible to exploit it online, to make inference about the intended goal. For the sake of simplicity, suppose that only the distribution related to one hand of the operator is updated. To this purpose, the measures retrieved from the sensor as well as the knowledge about every possible goal location are exploited to compute  $\Phi^1_{k+1}, \dots, \Phi^n_{k+1}$ , one for every goal. Then, Eq. (1) is adopted to compute the a posteriori probabilities of every goal at step  $k+1$ .

Since the information about  $z$  is not always valid, we consider a switching model for the update of probabilities. In case  $z$  is not available, we can make inference according to a reduced vector of features  $\Psi^i_k = (\theta^i_{W1k} \ d^i_k)^T$ . Another GMM distribution, let us call it GMM2d, can be learnt for  $\Psi$ , in a similar way as for the one describing  $\Phi$ , that can be denoted as GMM4d. Then, depending on the value returned at step  $k$  for flag  $f_k$ , the update of probabilities for every goal is made according to GMM4d, or according to GMM2d.

The overall pipeline of the presented approach is reported in Fig. 2.

#### E. Managing uncertainties for the location of the goals

As stated in the previous Section, the computation of the feature vector  $\Phi^i$  (or  $\Psi^i$ ) requires to know the position of the  $i^{th}$  goal. However, in realistic contexts, this quantity is not a precise value, but it is rather a random variable with a certain distribution of probability. To manage this uncertainty it is only required to characterize the distribution which describes the goal location. In this way, some possible locations for goals can be sampled every time the update of probability is required.

As done for particle filter algorithms, when computing the likelihood of the  $i^{th}$  uncertain goal we can consider a set of  $Np$  samples  $g^{i1}, \dots, g^{iNp}$  as hypothesis about the true goal location. Then, for every sample  $g^{ij}$ , it is possible to compute the likelihood function, by following the approach presented in Section II-D, which implies to compute different  $\Phi^{ij}$ . The global likelihood function adopted for updating the probabilities can be computed as a mean of the likelihood of every sample<sup>3</sup>:

$$\Pi(\theta_{k+1}|Goal_i) = \frac{1}{Np} \sum_{p=1}^{Np} \Pi(\theta_{k+1}|g^{ij}) = \frac{1}{Np} \sum_{p=1}^{Np} f(\Phi^{ij}).$$

### III. EXPERIMENTS

We tested our method in a collaborative task, where a human operator has to assemble a box with the help of a robotic manipulator. The experimental setup consists of an ABB dual-arm robot YUMI equipped with a suction cup and a Kinect depth camera to track human motions, see Fig. 3. The operator's left hand is equipped with a vibrotactile ring

(Fig. 4), which contains a 4 mm vibration motor (PRECISION MICRODRIVES<sup>TM</sup>)<sup>4</sup>, that is controlled through an Arduino Pro Mini<sup>5</sup>. The communication with the ring is wireless, thanks to two XBee<sup>®</sup> RF modules (Digi International Inc.)<sup>6</sup>. The ring itself weights around 2g, whereas the complete device (ring plus controller box placed on the wrist) weights around 40g.

The depth camera, the vibrotactile ring and the robot are connected to a CPU, where the inference algorithm is implemented. The CPU reads the measurements retrieved from the Kinect and sends commands to both the robot and the ring.

#### A. Description of the collaborative assembly task

The objective of the task is to assemble a box containing a USB pen drive. The task has 5 steps:

- 1) The human fills the box with two layers of foam, and the USB pen drive.
- 2) The human brings the filled box towards the robot.
- 3) The robot adds the cover on the box.
- 4) The human fixes the cover with some tape.
- 5) The robot stores the finished product.

The operator is monitored during the experiment by the depth camera, whose observations are given to the inference algorithm presented in Section II-D. The set of possible human goals is composed as follows (see Fig. 3):

- 1) Home position
- 2) Feeder of the first kind of layer foams
- 3) Feeder of the second kind of layer foams
- 4) Collaborative station
- 5) Feeder of the boxes and pen drives

The haptic feedback is used twice. The first time to inform the human that the robot understood his/her intention of putting the box in the shared workspace, *i.e.*, the collaborative station (beginning of step 2). The second time to inform the human that the robot understood that step 4 has been completed. The decision about the proper time to send one new feedback is made according to the evolution of probabilities about the human's goals, as well as according to the state machine depicted in Fig. 5. When at least one new human is detected in the scene, the state machine goes out of its initial state and reaches state 1. This state persists until the probability related to goal 4 grows above a predefined high threshold (0.8 for instance), meaning that the human has completed the first step and is about to deliver the partially assembled box to the collaborative station. When this happens, the machine enters state 2 and returns to state 1 only when the probability of goal 1 grows above another threshold, implying that the operator has finished step 4 and is about to begin a new cycle. Every time the machine goes from state 1 to state 2 or vice versa, a vibrotactile feedback (a vibration burst lasting 120 ms, with a frequency of 200 Hz and an amplitude of 0.8 g) is sent to the operator. We decided to use the same kind of feedback in both cases because the aim of the delivered

<sup>4</sup>Datasheet Motor

<sup>5</sup>Datasheet Arduino Pro Mini

<sup>6</sup>Datasheet XBee

<sup>3</sup>Which can translated in a weighted sum, in case samples are not equally probable.

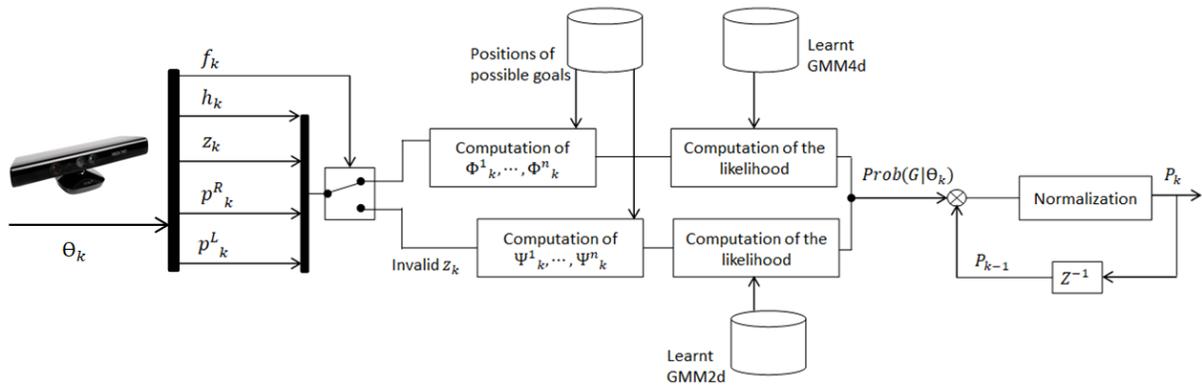


Fig. 2: Pipeline of the proposed approach to tackle the problem of inference. Flag  $f_k$  is exploited to select the proper GMM to compute the likelihood of every goal.

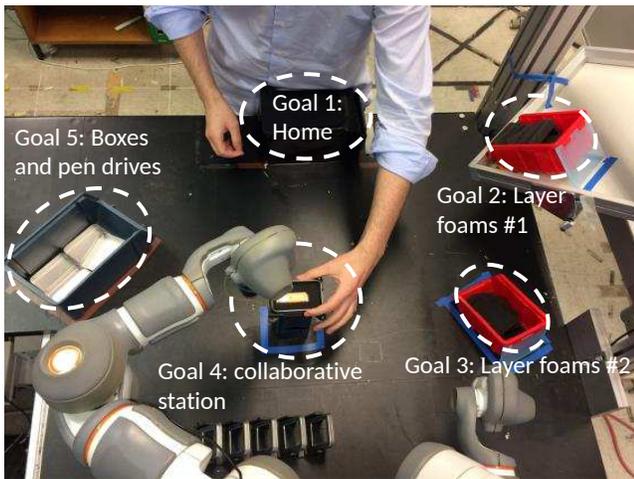


Fig. 3: Layout of the experimental robotic cell. Locations of possible human goals are indicated with white dotted circles.

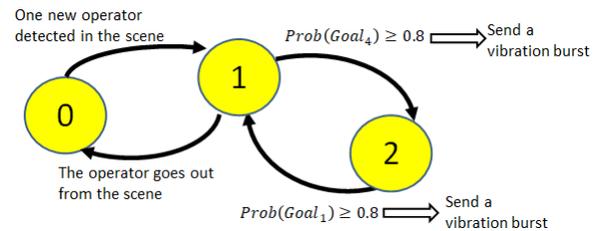


Fig. 5: State machine adopted to send feedback to the operator. Note that vibration bursts are sent before the operator actually reaches the goals, according to the probabilities evolution.

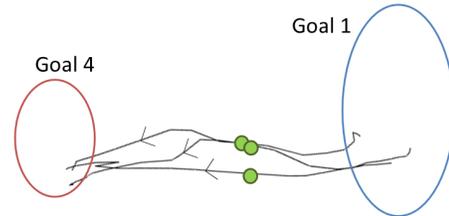


Fig. 6: Some trajectories taken from the experiments, going from goal 1 (blue) to goal 4 (red). Green markers are located at the points of the path for which the subject receives the haptic feedback.



Fig. 4: Vibrotactile ring with its controller box. During the experiments the ring is worn on the operator’s left hand and the box is attached to a Velcro bracelet worn on the forearm.

message is the same, *i.e.*, informing in a reactive way the human that the robot has understood his/her intention and is about to move.

### B. Experiments

We recruited 16 participants for our experiments. Half of them performed the collaborative task with the haptic ring, and half without. In both groups, 5 out of 8 subjects were considered “non-skilled”, as they declared to be not familiar

with the use of robots. The “skilled” participants, instead, had previous experience with human-robot collaborative tasks (*e.g.*, they already took part in HRC experiments), but were not specifically trained for the proposed assembly task. All subjects were asked to perform the collaborative task for 5 consecutive times, and those wearing the ring were instructed on the meaning of the vibration burst. During the experiments the ring was worn on the operator’s left hand, that was the one tracked with the Kinect. We recorded the execution time of each trial and we asked subjects that used the vibrotactile interface to evaluate their experience with a questionnaire. Fig. 6 shows some trajectories taken from one subject, with the point at which the haptic feedback is sent, which is almost in the middle of the path.

We adopted the method described in Section II-C to train GMM4d and GMM2d models. Results regarding training of GMM4d are reported in Fig. 7, where the likelihood of the model is plotted against the number of clusters considered.

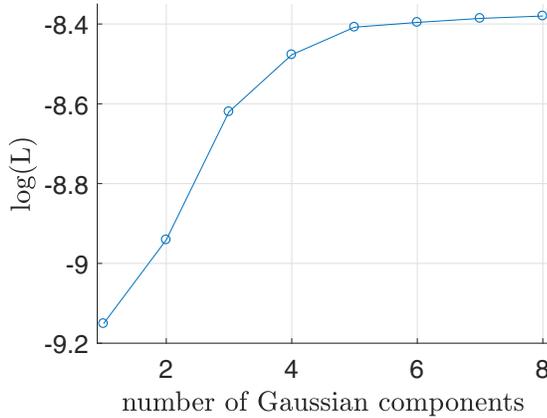


Fig. 7: The figure reports the likelihood of model GMM4d, varying the number of clusters considered when training the model. The selected number of clusters was 7.

A number of 7 clusters was considered for GMM4d since, as can be seen from Fig. 7, considering a greater number does not improve significantly the likelihood of the model. Experimental results are discussed in the next section.

#### IV. RESULTS AND DISCUSSION

We evaluated our collaborative framework analysing three different aspects: the benefits of using gaze estimation, the benefits of using the vibrotactile feedback, and the overall subjects’ evaluation of the haptic ring.

##### A. Benefits of using gaze estimation

To highlight the benefits of using the estimate of the gaze in the inference process, we compared the probabilities evolution obtained with three different methods:

- M1: the method proposed in [5], which considers in the inference process only the trajectory of a single hand of the operator;
- M2: an approach that updates the probabilities based on the trajectory of a single hand and using the likelihood function denoted as GMM2d, simulating that the information about gaze is always unavailable;
- M3: the complete approach proposed in Section II-D, where inference is made by considering both the detected positions of the operator’s hands and his/her gaze.

Method 3 was applied on-line, while the other two were applied off-line on the same measurements retrieved from the depth camera. Fig. 8 reports the distribution of the distance at which goal 4 was correctly recognized by using the aforementioned methods. As can be seen, the performance of the methods is quite similar. Anyway, the robustness of the three methods must be taken into account. We define the number of false positives as the number of times for which the probability of a certain goal has risen beyond the threshold of 0.8, but the operator was going to a different target. The number of true negatives, instead, corresponds to the number of times for which the operator was going to a certain goal, but the probability of the same did not rise above the threshold. From

the analysis of Table I, that reports data related to goals 1 and 4, we can state that the information about the gaze allows to acquire an improved robustness in the inference process. The position of goal 1 is much more scattered than goal 4, and this reflects on the performance of our algorithm, even though it remains better than the method proposed in [5].

TABLE I: Percentages of false positives and true negatives are with respect to the total number of times the operators went to goal 4 for the upper table. The lower one refers to goal 1.

Goal 4	M1	M2	M3*
% False Positives	4.54	11.53	<b>0</b>
% True Negatives	16.67	<b>0</b>	<b>0</b>
Goal 1	M1	M2	M3
% False Positives	9.09	11.15	<b>6.06</b>
% True Negatives	21.21	12.3	<b>8.3</b>

\*The third column of the first table contains zeros because during the experiments we did not notice neither false positives, nor true negatives (N.B. method 3 was applied on-line).

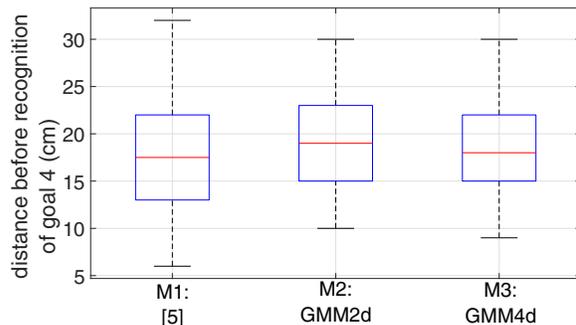


Fig. 8: Comparison of the three different inference methods regarding the goal recognition performance. Distance before recognition refers to the length of the path described by the wrist of the operator, from the instant at which the recognition of goal happens until the time at which the operator reaches goal 4.

##### B. Benefits of using haptic feedback

To evaluate the advantages of using haptic feedback, we decided to analyse the task execution time. We started from the intuition that if the human is aware that the robot has understood his/her intention, he/she does not have to wait to see it moving, and can proceed with the task execution in a more fluent and fast way. In other words, the haptic feedback makes the user confident that his/her artificial mate is working as expected.

From the data gathered in the experiments described in Section III, we observed that within the overall population, variability in cycle time is reduced with the help of vibrotactile feedback, however no statistical evidence of this fact can be proven (Fig. 9). Within the population of non-skilled participants (Fig. 10), instead, we found statistical evidence of the fact that the average cycle time is reduced with the help of vibrotactile feedback (single-tailed Wilcoxon rank sum test with confidence  $\alpha = 0.05$  returns  $r = 0.9907$ ). A tangible decrease of the cycle time variability for the same

population can also be appreciated, however still without statistical evidence.

We also considered the time elapsed from the instant the human finishes to fix the tape on the cover and the time he/she reaches goal 1 to begin a new assembling cycle, for both skilled and non-skilled participants (Fig. 11). The average time when receiving the vibrotactile feedback is strongly proved to be statistically lower (single-tailed Wilcoxon rank sum test with confidence  $\alpha = 0.05$  returns  $r = 0.9998$ ). This confirms the hypothesis that the usage of vibrotactile feedback for some crucial parts of the interaction makes the user more confident about the robot behaviour, improving his/her productivity.

The task that we present in this paper is rather easy, and that is probably why performance of people with previous experience in robotics does not seem to benefit from haptic feedback. However, having seen the results obtained with non-skilled subjects, we expect that in contexts where the production lines change frequently (*e.g.*, agile manufacturing), haptic feedback could prove to be significantly beneficial.

To further assess its usability in industrial contexts, haptic feedback could be compared with auditory or visual cues, in case hearing and sight are not impaired during the task execution. However, the comparison between different types of feedback modalities was out of the scope of this work. Future developments could address this issue and also study how multimodal interaction [28] can be combined with intention prediction algorithms to perform HRC tasks.

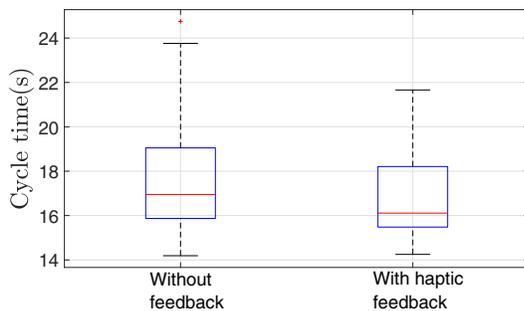


Fig. 9: Cycle time with and without ring. The overall population, composed of skilled and non-skilled subjects, was considered.

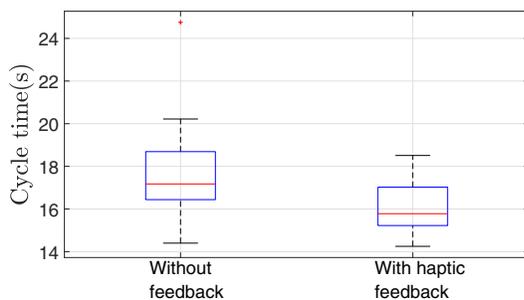


Fig. 10: Cycle time with and without ring for non-skilled subjects.

### C. Overall evaluation of the haptic ring

After the experiments done with the haptic ring, we submitted to the subjects a questionnaire to get their subjective

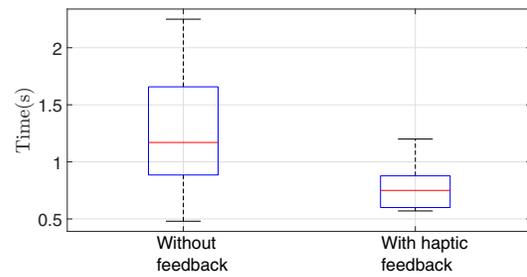
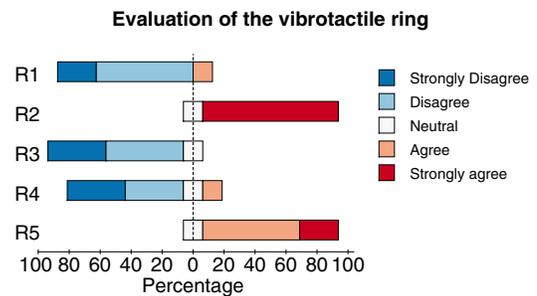


Fig. 11: Time elapsed from the instant when the tape is fixed until the one when the operator returns to goal 1. The overall population, composed of skilled and non-skilled subjects, was considered.

evaluation of the device. None of the 8 subjects had previously used wearable haptic interfaces before. We formulated the questions as five-level Likert items, and from the answers shown in Fig. 12 we can derive an overall appreciation of the ring, that was felt more as a help (cf. R5) than as an encumbrance (cf. R1, R3). Answers to question R5 are in line with the results shown in Fig. 11, underlining that the vibrotactile feedback helps in proceeding smoothly with the task.



- R1: I found the vibrotactile ring very cumbersome to use.
- R2: The vibrations produced by the ring are easy to distinguish.
- R3: Wearing the ring impedes some movements.
- R4: The vibration is annoying.
- R5: Having an acknowledgement from the robot helps in going on with the task.

Fig. 12: Subjects’ answers in percentage to the question “Express how much you agree with the following statements concerning the vibrotactile ring”.

In this paper, we focused on using a vibrotactile ring, as it is one of the most simple and unobtrusive wearable haptic devices [19], and acts directly on the user’s body part (the hand) that is mainly in charge of performing the assembly task. Other possible solutions, including different haptic devices (*e.g.*, bracelets, armbands, headbands), and different feedback signals (*e.g.*, vibrations patterns, vibration bursts with different durations, force feedback) could be tested in future works to convey meaningful information during HRC tasks.

## V. CONCLUSIONS

In this work, we propose a human-robot collaboration paradigm where human intentions are inferred through a Bayesian recursive classifier and are used for a twofold purpose. On

one side, to decide how and when the robot has to perform its part of the collaborative task, on the other, to send a vibrotactile feedback to the human operator to acknowledge that his/her intention has been recognized with high accuracy. Our inference algorithm forecasts, within a fixed and finite set of goal positions, the one that will be reached by the operator, relying on the on-line acquisition of the trajectories of the human hands and head. This algorithm was inspired by our previous work, that has been here extended to take into account the estimated human's gaze direction, which allows to achieve higher robustness to the inference process. Through experiments with 16 volunteers, we showed its applicability to a collaborative assembly task, and we found that adding haptic feedback to alert the human when the robot has understood her/his intentions, significantly improves the average execution time of non-skilled subjects. Future investigations will focus on studying how haptic feedback influences the learning of new tasks in naive users.

The combination of an intention prediction algorithm with haptic feedback is a new idea that could potentially be exploited in several human-robot collaborative tasks, choosing the inference algorithms and tactile displays that are most suitable for the particular application.

## REFERENCES

- [1] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human-robot collaboration," *Autonomous Robots*, pp. 1–19, 2017.
- [2] S. Robla-Gómez, V. M. Becerra, J. Llata, E. González-Sarabia, C. Torre-Ferrero, and J. Pérez-Oria, "Working together: a review on safe human-robot collaboration in industrial environments," *IEEE Access*, vol. 5, pp. 26 754–26 773, 2017.
- [3] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 14–29, 2016.
- [4] R. Luo, R. Hayne, and D. Berenson, "Unsupervised early prediction of human reaching for human-robot collaboration in shared workspaces," *Autonomous Robots*, pp. 1–18, 2017.
- [5] A. M. Zanchettin and P. Rocco, "Probabilistic inference of human arm reaching target for effective human-robot collaboration," in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. IEEE, 2017.
- [6] H. C. Ravichandar and A. P. Dani, "Human intention inference using expectation-maximization algorithm with online model learning," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 855–868, 2017.
- [7] J. L. Drury, J. Scholtz, and H. A. Yanco, "Awareness in human-robot interactions," in *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, vol. 1. IEEE, 2003, pp. 912–918.
- [8] E. K. Phillips and F. G. Jentsch, "Supporting situation awareness through robot-to-human information exchanges under conditions of visuospatial perspective taking," *Journal of Human-Robot Interaction*, vol. 6, no. 3, pp. 92–117, 2017.
- [9] S. Makris, P. Karagiannis, S. Koukas, and A.-S. Matthaiakis, "Augmented reality system for operator support in human-robot collaborative assembly," *CIRP Annals-Manufacturing Technology*, vol. 65, no. 1, pp. 61–64, 2016.
- [10] T. Johnson, G. Tang, S. R. Fletcher, and P. Webb, "Investigating the effects of signal light position on human workload and reaction time in human-robot collaboration tasks," in *Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future*. Springer, 2016, pp. 207–215.
- [11] A. St Clair and M. Mataric, "How robot verbal feedback can improve team performance in human-robot task collaborations," in *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*. ACM, 2015, pp. 213–220.
- [12] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, 2018.
- [13] L. A. Jones and N. B. Sarter, "Tactile displays: Guidance for their design and application," *Human Factors*, vol. 50, no. 1, pp. 90–111, 2008.
- [14] D. Prattichizzo, F. Chinello, C. Pacchierotti, and M. Malvezzi, "Towards wearability in fingertip haptics: a 3-dof wearable device for cutaneous force feedback," *IEEE Transactions on Haptics*, vol. 6, no. 4, pp. 506–516, 2013.
- [15] C. Pacchierotti, S. Sinclair, M. Solazzi, A. Frisoli, V. Hayward, and D. Prattichizzo, "Wearable haptic systems for the fingertip and the hand: Taxonomy, review, and perspectives," *IEEE Transactions on Haptics*, vol. 10, no. 4, pp. 580–600, Oct 2017.
- [16] S. Musić, G. Salvietti, P. B. gen. Dohmann, F. Chinello, D. Prattichizzo, and S. Hirche, "Robot team teleoperation for cooperative manipulation using wearable haptics," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2017, pp. 2556–2563.
- [17] S. Scheggi, M. Agravi, F. Morbidi, and D. Prattichizzo, "Cooperative human-robot haptic navigation," in *Proc. IEEE Int. Conf. on Robotics and Automation*, no. 0, Hong Kong, China, 2014, pp. 2693–2698.
- [18] M. Maisto, C. Pacchierotti, F. Chinello, G. Salvietti, A. D. Luca, and D. Prattichizzo, "Evaluation of wearable haptic systems for the fingers in augmented reality applications," *IEEE Transactions on Haptics*, vol. 10, no. 4, pp. 511–522, Oct 2017.
- [19] I. Hussain, L. Meli, C. Pacchierotti, G. Salvietti, and D. Prattichizzo, "Vibrotactile haptic feedback for intuitive control of robotic extra fingers," in *2015 IEEE World Haptics Conference (WHC)*, June 2015, pp. 394–399.
- [20] I. Hussain, L. Meli, C. Pacchierotti, and D. Prattichizzo, "A soft robotic supernumerary finger and a wearable cutaneous finger interface to compensate the missing grasping capabilities in chronic stroke patients," in *2017 IEEE World Haptics Conference (WHC)*, June 2017, pp. 183–188.
- [21] V. Hayward and K. E. Maclean, "Do it yourself haptics: part i," *IEEE Robotics Automation Magazine*, vol. 14, no. 4, pp. 88–104, Dec 2007.
- [22] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 951–970, 2013.
- [23] Z. Zhang, A. Beck, and N. Magnenat-Thalmann, "Human-like behavior generation based on head-arms model for robot tracking external targets and body parts," *IEEE transactions on cybernetics*, vol. 45, no. 8, pp. 1390–1400, 2015.
- [24] C. L. Kleinke, "Gaze and eye contact: a research review," *Psychological bulletin*, vol. 100, no. 1, p. 78, 1986.
- [25] R. Bednarik, H. Vrzakova, and M. Hradis, "What do you want to do next: a novel approach for intent prediction in gaze-based interaction," in *Proceedings of the symposium on eye tracking research and applications*. ACM, 2012, pp. 83–90.
- [26] M. Carrasco and X. Clady, "Prediction of user's grasping intentions based on eye-hand coordination," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 4631–4637.
- [27] J. A. Bilmes *et al.*, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," *International Computer Science Institute*, vol. 4, no. 510, p. 126, 1998.
- [28] K. S. Hale and K. M. Stanney, "Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations," *IEEE Computer Graphics and Applications*, vol. 24, no. 2, pp. 33–39, March 2004.