Affect Recognition in Hand-Object Interaction Using Object-sensed Tactile and Kinematic Data

Radoslaw Niewiadomski, Cigdem Beyan and Alessandra Sciutti

***Abstract*—We investigate the recognition of the affective states of a person performing an action with an object, by processing the object-sensed data. We focus on sequences of basic actions such as grasping and rotating, which are constituents of daily-life interactions. iCube, a 5 cm cube, was used to collect tactile and kinematics data that consist of tactile maps (without information on the pressure applied to the surface), and rotations. We conduct two studies: classification of *i)* emotions and *ii)* the vitality forms. In both, the participants perform a semi-structured task composed of basic actions. For emotion recognition, 237 trials by 11 participants associated with anger, sadness, excitement, and gratitude were used to train models using 10 hand-crafted features. The classifier accuracy reaches up to 82.7%. Interestingly, the same classifier when learned exclusively with the tactile data performs on par with its counterpart modeled with all 10 features. For the second study, 1135 trials by 10 participants were used to classify two vitality forms. The best-performing model differentiated gentle actions from rude ones with an accuracy of 84.85%. The results also confirm that people touch objects differently when performing these basic actions with different affective states and attitudes.**

***Index Terms*—affective touch; emotion classification; hand-object interaction; vitality forms; tactile data**

I. Introduction

In our daily life, we often grasp, move or pass some small objects, such as boxes or mugs with our hand(s). In our research, we aim to recognize the internal states from the data collected during such *natural* hand-object interactions (HOI). *Natural interactions* in our scope refer to performing basic actions such as grasping, rotating, and holding (as well as combinations of them), which are elements of the daily interactions with a variety of objects [1]. In this line, Wang et al. [2] discuss *non-symbolic touch* as "no predefined meaning or code is needed for affect conveyance using touch". Such non-symbolic touch may appear in human-human interactions (HHI) (e.g., a child squeezes her/his mother's hand when being afraid on a roller coaster ride). That is more frequent when manipulating some objects (e.g., one squeezes a mug or a pen when being upset). It is important to distinguish non-symbolic touch from the often conventional symbolic action (i.e., social touch) performed explicitly to communicate some affective states, and interpersonal attitudes such as handshake, hug, finger-interlocking, or hand-lifting. When grasping or rotating an object with one or two hands, the information about the human's affective state might potentially be inferred from multiple sensors placed on/inside of the object resulting in *object-sensed* data, as well as from the sensors placed on the human body resulting in *agent-sensed* data. In this paper, we focus on object-sensed data only.

There are at least three important motivations behind this work. First, using tactile data from sensors placed on the object, and in particular, the tactile maps, with no information about the pressure applied to the object, is a new and innovative approach to the well-known problem of affective state recognition. The benefits of using such types of data include lower complexity, potentially lower costs, and faster data processing. Moreover, this method could be relatively easily integrated into the objects while bringing in new perspectives to several applications in affective computing. Second, our approach complements the research on affect recognition with *agent-based* data, e.g., the works using wearable gloves to collect the tactile data on affective hand-object interaction. In detail, our *object-sensed* approach introduces several benefits such that the user does not need to be endowed in any specific hardware to start an interaction while when using the glove, she is always aware of the carried sensing device. Wearing it may be considered intrusive, and cumbersome, and as a consequence, may have an impact on (affective) interaction. On the other hand, when the object-sensed data approach is applied, the users might interact more naturally even without being aware of the process of tactile data collection. Additionally, *object-sensed* approach can be easily extended to social settings (i.e., two or more people interact with the same object). Third, this research addresses the question of whether or not people touch objects differently when performing basic actions such as rotating or grasping with different affective states.

Previous works on social and affective touch in HHI (e.g., [3], [4]) and human-robot interactions (HRI) (e.g., [5], [6]) often considered conventional gestures performed upon explicit request of the experimenter. Some studies addressed social touch recognition. For example, in [7], [8] several actions such as "grasping the arm suddenly" (grab) and "rubbing the arm with the hands" (massage) were classified using the data collected with a tactile grid placed on a mannequin forearm. Instead, the works on non-symbolic gestures often used *soft* objects such as balloons, and analyzed the data on the pressure applied during the contact [9]–[11]. Others focused on multi-touch screens when humans play video games [12] or tap virtual keyboards [13]. There also exists an approach based on measuring the directional photoreflectivity [14].

Unlike the aforementioned works, we focus on non-

R. Niewiadomski is with the Department of Psychology and Cognitive Science, University of Trento, Rovereto, Italy. E-mail: r.niewiadomski@unitn.it.

C. Beyan is with the Department of Information Engineering and Computer, University of Trento, Trento, Italy, e-mail: cigdem.beyan@unitn.it.

A. Sciutti is with the CONTACT Unit, Istituto Italiano di Tecnologia, via Melen 83, Genoa, Italy, e-mail: alessandra.sciutti@iit.it.

symbolic touch actions during the HOI, which can, but does not have to, be performed in a social context, i.e., in the presence of another person. This study is the continuation of our prior research [15] which considered *only emotion recognition* during HOI by presenting a preliminary model with *limited* number of features. Instead in this paper, we investigate whether the affective states, i.e., different **emotions** and **vitality forms**, can be differentiated using hand-crafted features extracted from tactile and kinematics data. Whilst emotions are short-lasting responses to relevant events that involve synchronized changes in multiple subsystems [16], vitality forms reflect the performer's attitudes and they may modulate human nonverbal behavior in a continuous manner. By considering different affective phenomena, we aim to show that our approach is *generalizable* and can be applied to various contexts and action combinations. At the same time, we acknowledge that the way one interacts with an object depends on certain properties of that object (e.g., its dimension, shape, hardness) as well as the activity performed. In this work, we control such factors by choosing one specific semantically-neutral object, and by restricting the number of activities. Keeping above mentioned factors fixed, we focus on how our internal states influence the way we touch objects.

II. iCube device

iCube version 2.0 [17], [18] is a 5 cm wireless hard cube weighing about 150 grams (see Figure 1) created at Istituto Italiano di Tecnologia.[1] It generates an asynchronous combination of tactile (i.e., 2D tactile maps) and kinematics (i.e., angle rotations in quaternions) data. The sampling rate is around 10 samples per second for the tactile data. Each face of the cube is covered by the $4 \times 4 = 16$ Capacitive Button Controllers that are able to detect simultaneous touches. The reader can refer to [17] for other technical details. The touch pressure data is not collected by the device, and this choice allows us to reduce the cost of the device, and increase the amount of data that could be sent to the computer in real time.

The main advantage of using the iCube to collect affect-related data is its semantically-neutral and simple shape. iCube is similar, in terms of shape and weight, to several objects (e.g., small containers) that a person interacts with in daily life. We believe that such similarities could allow our participants to carry out natural interactions with the cube, by allowing to exploit "users' pre-existing understanding and interaction with similar objects from their everyday world" [10].

III. Features

We propose a total of 10 features. Seven of these are extracted from the tactile data, two represent the kinematic property, and the last one is related to the action duration. When designing features, we took inspiration from the prior HHI and HRI studies. For instance, Wallbott [19] and Castellano et al. [20] have shown that humans perform more *expansive* and *quick* gestures when they feel high arousal emotions such as anger, whilst they may tend to slow down the same gestures when feeling sad. Masson et al. [4] observed a positive correlation between the rated arousal and the *motion energy* during interpersonal socio-affective touch actions. A study on HRI [5] shows that attachment emotions (e.g., gratitude, sadness) are characterized by *longer tactile contact* than the rejective ones (e.g., anger, disgust). In [21], emotions are differentiated during the hand-forearm contact using *the total contact area*, *the touch duration*, and the hand velocity. Consequently, our features measure the task duration, the amount of movement, the area of physical contact, and the touch variability.

[1]A video of an interaction with iCube is included as a Supplementary Material.

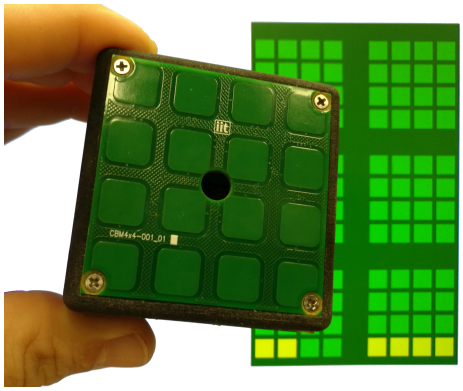

Fig. 1. iCube device. The real-time tactile data visualization can be seen in the background.

A. Tactile data

Let $a_{ijkm} = 1$ if a cell on the intersection of i-th row and j-th column of the k-th face (pad) of iCube is touched at the moment (i.e., single readout of all tactile data) m of a data segment; and $a_{ijkm} = 0$ if the same cell is not touched. Let p_{km} be a k-th face, $k = 1...6$, at the moment m, $m = 1...n$, where n corresponds to the total length (i.e., the number of readouts) of a data segment.

Density. The touch density estimates how large the portion of the surface of the cube is engaged in contact. We compute an average number of touched cells in a data segment (aTD) as:

$$aTD = \frac{\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{6}\sum_{m=1}^{n} a_{ijkm}}{n} \quad (1)$$

The higher the aTD is, the larger surface of the cube is touched on average. We also compute the maximal value (mTD) on the whole data segment.

Variability. The touch variability is used to estimate the quantity of contact changes during the task. We compute the number of changes in touched cells between two consecutive readouts, and then we compute the average value for a data segment. We introduce average touch variability (aTV) as:

$$aTV = \frac{\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{6}\sum_{m=2}^{n} |a_{ijkm} - a_{ijkm-1}|}{n-1} \quad (2)$$

We also compute the maximal value (mTV) of the touch variability on the data segment.

Allocation. We count how often the central and the peripheral cells are activated as well as whether the touch concerns a whole pad or it is localized in one (or more) pad quarters. First, we check if the central cells of k-th pad at the moment m are touched:

$$cent_{km} = \begin{cases} 1 \text{ if } a_{ijkm} = 1 \text{ for any pair } i,j: \\ \quad \{2,2\}, \{2,3\}, \{3,2\}, \{3,3\}, \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

Next, we compute the average number of the *central touches* (aCT) as:

$$aCT = \sum_{k=1}^{6} \sum_{m=1}^{n} \frac{cent_{km}}{n} \tag{4}$$

Similarly, we compute $side_{km}$ by considering the following indices pairs: $\{1,1\}$, $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,1\}$, $\{3,1\}$, $\{2,4\}$, $\{3,4\}$, $\{4,1\}$, $\{4,2\}$, $\{4,3\}$, $\{4,4\}$. Finally, by analogy to aCT case, we compute the average number of *peripheral touches* (aPT) using values of $side_{km}$. We also measure the touch dispersion on a pad p_{km} according to the following procedure:

$disp_{k,m} \leftarrow 0$
if $a_{1,1,k,m} = 1$ **or** $a_{1,2,k,m} = 1$ **or** $a_{1,3,k,m} = 1$ **or** $a_{1,4,k,m} = 1$ **then**
  $disp_{k,m} \leftarrow disp_{k,m} + 0.25$
end if
if $a_{1,1,k,m} = 1$ **or** $a_{1,2,k,m} = 1$ **or** $a_{1,3,k,m} = 1$ **or** $a_{1,4,k,m} = 1$ **then**
  $disp_{k,m} \leftarrow disp_{k,m} + 0.25$
end if
if $a_{1,1,k,m} = 1$ **or** $a_{1,2,k,m} = 1$ **or** $a_{1,3,k,m} = 1$ **or** $a_{1,4,k,m} = 1$ **then**
  $disp_{k,m} \leftarrow disp_{k,m} + 0.25$
end if
if $a_{1,1,k,m} = 1$ **or** $a_{1,2,k,m} = 1$ **or** $a_{1,3,k,m} = 1$ **or** $a_{1,4,k,m} = 1$ **then**
  $disp_{k,m} \leftarrow disp_{k,m} + 0.25$
end if

The values of $disp_{km}$ are in the interval of $[0,1]$ when the higher values mean that the touch is more spread over the pad. Next, we compute the average dispersion (aDT) as:

$$aDT = \sum_{k=1}^{6} \sum_{m=1}^{n} \frac{disp_{km}}{n} \tag{5}$$

B. Kinematics data and Duration

Rotation. To estimate the movement quantity, we compute the total number of rotations. More specifically, we compute the instantaneous angular variation by measuring the angle traversed over the time for each of the three unitary axes orthogonal to the faces of the cube using the method described in [18]. To quantify the total amount of rotation, we compute the maximum value among three cumulative sums of the rotations. Then, we compute the average (aTR), and the maximum value (mTR) on the data of the mostly rotated axis over the whole segment.

We also consider the total duration, referred to as $TIME$, of a data segment in seconds.

IV. Study 1: Emotion Classification

Data Collection. The task consists of taking the iCube placed on the table, finding a marker that is attached to one of its pads, and passing the cube to another person (so-called confederate) in a way that he/she can see the marker. Performing this task requires basic actions such as grasping, rotating, and handing over the iCube. Our participants were asked to *imagine* to feel a specific emotion when performing the task. These emotions are anger, sadness, excitement, and gratitude. They are placed in four different quadrants of the two-dimensional valence-arousal model. The first three are mentioned in Russell's paper [22] such that the anger and the excitement are characterized by high arousal, whilst the anger and the sadness are with negative valence. The gratitude does not appear in [22], but it was evaluated in later works. E.g., in [23] gratitude is positive, but the sixth lowest arousal emotion out of 62 labels, receiving a score of three on a nine-point scale. Thus, it is reasonable to assume that this positive state is characterized by lower arousal compared to excitement. A similar approach, i.e., four distinct labels corresponding to four quadrants of the valence-arousal model, was used in previous works, e.g., [12] to address touch-based interaction.

For the data collection, two assignments were designed. Each participant performed only one of them, chosen randomly. In the first assignment (A1), short written stories were used to provide emotional context to the participants. The task is fixed, but the emotion and the imaginary object mentioned in the story are different. In detail, in the case of gratitude, iCube becomes a small box of chocolates to be offered as a gift for a favor received. For the sadness scenario, the participants grab and pass a broken beloved wooden figure. In the case of anger, the cube becomes an empty packet passed to the confederate while accusing him of stealing its content. For the excitement scenario, the cube becomes an unexpected closed parcel addressed to the participant, who when passing it to the confederate asks to unbox it. The task is the same because the participants always need to *a)* grasp the iCube, *b)* rotate it to find the marker placed on one of the faces of the cube, *c)* approach the confederate, and *d)* pass iCube to the confederate. When defining the scenarios, we paid attention to introducing the "imaginary" objects (e.g., a small box of chocolates) that would match the iCube dimensions. The scenarios were written on different paper sheets, and their order of them was randomized. The participants picked one scenario at a time. We gave them some time to think about the story and to imagine themselves being the protagonist. Each participant performed 5-6 trials, and there was a 5-10 seconds pause between trials.

For the second assignment (A2), the participants were instructed to perform the same task portraying the four above-mentioned emotions. Unlike A1, in A2, the instructions regarding what the imaginary object and the scenario could be, were not given to the participants. In other words, in A2, the

paper sheets contained only the emotion labels. Before the data collection, the definitions of the four emotions, taken from [24], [25], were given to the participants.

For both A1 and A2, the initial positions of the participants, confederate, and the tables were kept always the same. When a participant was facing the confederate, the iCube was placed on a table which was on the left side of the participant, and the paper sheets were placed on another table placed on the participant's right. The confederate's position was fixed about 3-4 meters away from the participant's initial position. The iCube had a sticker on one of the pads, which symbolizes the front of the imaginary object (e.g., the opening of a box). The cube was placed in a way that participants cannot see the marker at the beginning of a trial. Given the portrayed emotion, the participants were asked to perform the task in the most natural way for them. The task is semi-structured, i.e., only a high-level description was given, but not the details (e.g., how to rotate the cube to find a marker, how to approach the confederate, whether or not to turn towards the cube, etc). We also intentionally avoid imposing the participants to perform an exact number of rotations, or to grab the cube in a fixed manner. This choice was made because we believe that the way the person performs the basic actions contains the affective information. For example, a person might rotate the object more or faster when she is angry compared to when she is sad. Giving more precise instructions would, in our view, impede our participants to behave naturally. Additionally, each participant performs the task in a different personal manner, which allows us to obtain interesting variability in the data. This would not be possible if we gave the participants exact and fixed instructions regarding the basic actions needed to accomplish the task.

The following classification experiments were conducted when the data of A1 and A2 were merged. By merging them as a single dataset, we expect that the classifiers' robustness would be improved. The data collection was performed with 11 participants (8 female, 1 left-handed). This resulted in 237 trials; composed of 60 sad, 59 angry, 59 excited, and 59 grateful data. From each trial, we extracted one data segment. A *data segment* in this study corresponds to the data captured from the time a participant makes physical contact with the iCube for the first time until she hands over the iCube. The average segment length is 3.9 seconds ($SD = 1.48s$).

Statistical Analysis. To examine the statistical differences between the four classes, we performed a series of Kruskal-Wallis tests with Emotion as the independent variable, each feature as the dependent variable, and by considering each data segment separately. A significant main effect of Emotion ($F(3, 228) = 3.665$, $p < 0.001$) on the segment duration (the variable $TIME$) was observed. Post-hoc comparisons using the Dunn-Bonferroni test showed that the segments with sad labels ($mean = 4.83s$) were significantly longer than angry ($p < 0.001$), excited ($p < 0.001$), and grateful segments ($p < 0.005$), whilst the grateful performances ($mean = 4.2s$) were longer than angry ($mean = 3.13s$, $p < 0.001$) and excited ones ($mean = 3.56s$, $p < 0.01$). The significant results were also observed for the average (aTD) and maximum (mTD) of touch density; the average touch variability (aTV); the average (aTR) and maximum rotation (mTR); as well as aPT and aDT. On average, a larger surface of the iCube was contacted for anger and excitement compared to sadness. At the same time, less touch variability was observed for sadness as compared to anger and excitement. More rotations were performed for the two high-arousal emotions: anger and excitement, as compared to sadness and gratitude.

Classification. We explored the performance of (a) Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel and (b) Localized Multiple Kernel Learning (LMKL) [26]. SVM-RBF was chosen as it was widely used to classify the emotions from the tactile [7], [11] and kinematics data [27]–[29]. LMKL showed significantly better performance for many applications involving human nonverbal behaviors analysis (e.g., [30]–[32]).

For each experiment, we performed *i)* leave-one-trial-out (LOTO) and *ii)* leave-one-subject-out (LOSO) cross-validation (CV) techniques when 10-features (described in Sec. III) and only 7 tactile features (aTD, mTD, aTV, mTV, aCT, aPT, aDT) were used. LOTO can be described as: given a dataset composed of N samples, $N-1$ samples are used for training (and validation) while the corresponding test split has only one sample that does not overlap with the training data. Such training-testing procedure is repeated for N times. On the other hand, LOSO can be defined as: given a dataset composed of N samples, the test split is composed of M samples belonging to one subject, and the corresponding training split is composed of $N - M$ data belonging to the other subjects. Such training-testing procedure is repeated as much as the number of subjects. It is important to mention that we also used validation data, which was picked randomly from the training splits and its size was taken as the 10% of the training data. Following the common practice, we used the validation data to select the parameters of SVM-RBF and LMKL.

The kernel parameters of SVM-RBF namely, C (which is a trade-off parameter between model simplicity and classification error) and γ were taken as the consecutive odd powers of two in the range of $[-7, 15]$. On the other hand, LMKL uses a nonlinear kernel weights combination. It has two components, called *i)* the gating model, which selects the optimum kernel function locally, and *ii)* the kernel-based classifier (which is SVM in this paper for a fair comparison). As the gating model softmax function was used with linear kernels varying from two to five. As kernel parameter C was taken as the consecutive odd powers of two in the range $[-11, 11]$.

The results are given in Table I. For all cases, all performance metrics are highly above the random chance level (around 25%). LMKL always performed much better than SVM-RBF, which is consistent with the previous works that use both methods [30]–[32]. As expected (due to having more training data as well as not considering the interpersonal differences) the results obtained with LOTO are higher than the LOSO counterpart. Still, it is important to notice that the LMKL results for LOSO are very remarkable (i.e., F-score of 77.79% and 77.60% for 10 and 7 features, re-

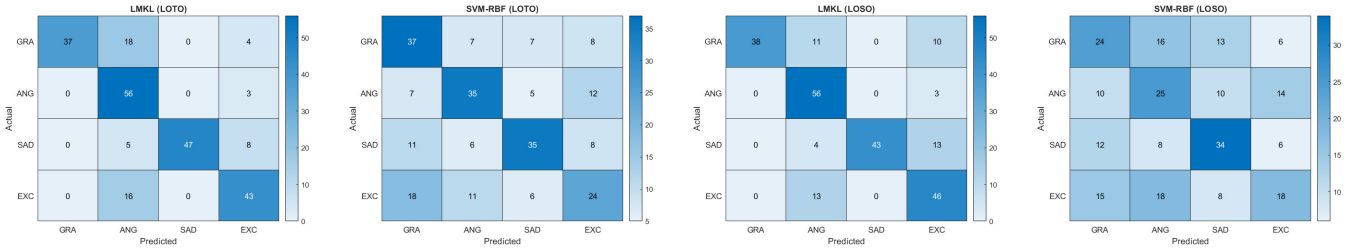

Fig. 2. Confusion matrix of LMKL and SVM-RBF for the classification of angry (ANG), sad, excited (EXC), and grateful (GRA) when 10 features are used and LOTO and LOSO are applied.

TABLE I
EMOTION CLASSIFICATION RESULTS (%) OBTAINED BY APPLYING LOTO (FIRST) / LOSO (SECOND). ALL METRICS ARE IN TERMS OF MACRO AVERAGE. THE BEST RESULTS ARE SHOWN IN BOLD.

	Accuracy	F-score	Precision	Recall
SVM-RBF	55.27 / 42.62	55.14 / 42.26	55.55 / 42.47	55.26 / 42.56
SVM-RBF (tactile only)	45.15 / 37.98	45.40 / 37.99	46.05 / 38.36	45.15 / 37.94
LMKL	**82.70 / 77.22**	**83.01 / 77.79**	**87.36 / 83.27**	**82.71** / 77.21
LMKL (tactile only)	**82.70 / 77.22**	82.96 / 77.60	87.16 / 82.64	82.68 / **77.24**

spectively), showing that LMKL is able to generalize well across subjects. For SVM-RBF, the results obtained with 7 features are always highly above the chance level but they are noticeably lower than the performances obtained with 10 features. Instead, LMKL's performance with 7 features is only slightly lower than the results with 10 features for average macro F-score, precision, and recall. This shows that LMKL is better than SVM-RBF to determine the kernel weights, and as shown in [31], this fact can be used to determine the best-performing features. Following [31], we calculated the average absolute kernel weights per feature obtained from LMKL. The important features should have higher combination weights. According to this rule, overall the best performing five features respectively are mTV, aPT, aTV, aDT, aCT. The best performances of SVM-RBF with 10 features were obtained when LOTO was applied with $C = 7$, $\gamma = -5$, and LOSO was applied with $C = 5$, $\gamma = -5$. Using 7 tactile features, the best performances of SVM-RBF were obtained when LOTO was applied with $C = 11$, $\gamma = -7$, and LOSO was applied with $C = 9$, $\gamma = -7$. The best performance of LMKL with 10 features was obtained when LOTO was applied with $C = 5$, the number of kernels 3, and LOSO was applied with $C = 1$, the number of kernels 2. With 7 features, LMKL's best performances were obtained when LOTO was applied with $C = 7$ and 3 kernels, and LOSO was applied with $C = -11$ and 5 kernels.

Figure 2 shows the confusion matrices of LMKL and SVM-RBF for the classification of emotions when 10 features are used and LOTO and LOSO are applied. As seen, for LKML, overall the best performances were obtained for anger. Sadness, the second negative emotion, was sometimes misclassified as anger or excitement. The worst recognition was observed for gratitude, which was sometimes misclassified as anger or excitement. For SVM, the best results were obtained for the sadness. The sadness was sometimes misclassified as gratitude (please notice that they have both low arousal). On the other hand, anger was sometimes misclassified as excitement, which is the second high-arousal emotion. On average, the worst recognition results were obtained for excitement, which was mainly misclassified as gratitude.

V. STUDY 2: VITALITY FORMS CLASSIFICATION

Humans may execute the same action in different ways such as vigorously, gently, or rudely. Vitality forms [33]; "how an action is performed", conveys important information about the performer's attitude. It reflects the internal states of the performer, providing an appraisal of the affective quality underlying the relation between him and the interaction partner(s). In this second study, we investigate whether the object-sensed tactile and kinematics data can be used to classify the vitality forms.

Data Collection. We asked 10 participants to perform a similar task presented in Study 1, i.e., taking the iCube placed on the table, finding a marker that is attached to one of its pads, and passing the iCube to the confederate in a way that she/he can see the marker. While performing this task, we asked participants to display two vitality forms (i.e., attitudes towards the confederate). These two vitality forms, namely rude and gentle, are the ones that were widely studied in the past. For example, by analyzing fMRI data, it was demonstrated that there exist different responses to these two vitality forms when the stimuli are visual [34] (i.e., gestures) and tactile (i.e., social touch) [35], meaning that these two vitality forms can be effectively transmitted through the aforementioned modalities. Therefore, it is reasonable to expect that object-sensed kinematic and tactile data can also be used to distinguish them.

Before the data collection, the short videos of other single-hand actions performed with two vitality forms, which were pre-validated in the previous fMRI study, were shown to the participants. During the presentation of these videos, we intentionally avoided pronouncing the labels identifying the vitality forms. The participants were asked to focus on the attitudes of the individuals in the videos (aka internal states) and not only to pay attention to the kinematic properties of the demonstrated actions. This is because the participants' task was to reproduce the same attitudes while the actions can be different from what they watched. Finally, the participants performed the task in the presence of a confederate, expressing the two attitudes observed previously.

The positions of the participants and iCube were as follows. When a participant was facing the confederate, the iCube was placed on a table that was on the participant's left side. The participants were sitting in front of the confederate (referred to as settings one and two) or they were standing about 3-4 meters away from the confederate's initial position (referred to as settings three and four). In all settings, there was a table between the participant and the confederate. The iCube had a sticker on one of the pads that participants could not see at the beginning of a trial. The participants were asked to perform the task in the most natural way given the vitality form they want to transmit. We did not introduce any additional constraints on how to perform constituent actions.

To introduce more variability to the data, all participants performed the task in four different settings. In the first two settings, while the participants were sitting in front of the confederate, *a)* they were supposed to take the iCube, and hand it to the confederate in a way that the marker is oriented towards the confederate or *b)* to take the iCube, and put it on the table in front of the confederate in a way that the marker is oriented towards the confederate. In the remaining two settings, the participants performed the task while standing. In the third setting, participants hand the cube to the confederate in the air (as in the first setting), and, in the fourth, they put the iCube on the table (as in the second setting). In particular, setting three corresponds to the setup of Study 1. Additionally, three different positions of the confederate were used during the data collection. First, the confederate was sitting exactly in front of the participant; second, she was placed at the right corner of the table, and, third, she was placed at the left corner of the table. For each setting and the confederate's position, the task was repeated at least five times. Due to technical issues, some trials of two participants (in setting 2 and 4) were not captured correctly and were discarded from the analysis. Consequently, the final dataset contains 1135 trials by ten participants.

Classification. The same methodologies (SVM-RBF and LMKL) and the cross-validation procedures described in Sec. IV were also applied for the classification of the vitality forms. The corresponding results are reported in Table II. The performances of SVM-RBF and LMKL are both higher than 83% for all metrics when 10 features are used. The results decrease (around 5-7%) when only tactile features are used. Similar to the results of Study 1, LOTO results are higher than LOSO up to 4%, showing that there exist some interpersonal differences. Nevertheless, these results are much higher than the random chance level.

For all cases, LMKL performs better than SVM-RBF. Using 10 features, the best performances of SVM-RBF were obtained when LOTO was applied with $C = 1$, $\gamma = -3$, and LOSO was applied with $C = -1$, $\gamma = -3$. Using 7 tactile features, the best performances of SVM-RBF were obtained when LOTO was applied with $C = 11$, $\gamma = -7$, and LOSO was applied with $C = 5$, $\gamma = -7$. The best performances of LMKL with 10 features was obtained when LOTO was applied with $C = -11$, the number of kernels$= 3$, and LOSO was applied with $C = -11$, the number of kernels$= 3$. LMKL's best performances using 7 features, were obtained when LOTO was applied with $C = 11$, the number of kernels$= 3$, and LOSO was applied with $C = 7$, the number of kernels$= 3$. Overall, the best performing five features among all are: aTD, aCT, TIME, aTV and aDT. Four out of five are tactile features, showing the importance of them with respect to remaining features, while, as can be noticed from the quantitative results, including kinematics features and the duration further improve the results.

TABLE II
VITALITY FORMS CLASSIFICATION RESULTS (%) OBTAINED BY APPLYING LOTO (FIRST) / LOSO (SECOND). ALL METRICS ARE IN TERMS OF MACRO AVERAGE. THE BEST RESULTS ARE SHOWN IN BOLD.

	Accuracy	F-score	Precision	Recall
SVM-RBF	83.79 / 80.62	83.79 / 80.54	83.80 / 81.17	83.79 / 80.63
SVM-RBF (tactile only)	75.60 / 73.74	75.60 / 73.74	75.61 / 73.75	75.60 / 73.74
LMKL	**84.85 / 81.32**	**84.80 / 81.32**	**85.35 / 81.34**	**84.86 / 81.33**
LMKL (tactile only)	76.48 / 74.36	76.48 / 74.35	76.49 / 74.40	76.48 / 74.36

Figure 3 shows the corresponding confusion matrices of LMKL and SVM-RBF for the classification of vitality forms when 10 features are used and LOTO and LOSO are applied. Except LMKL LOSO, in all cases, one can observe that the class accuracy of gentle is slightly higher than the class accuracy of rude.

VI. CONCLUSIONS

We showed that it is possible to recognize emotions and vitality forms during daily life non-symbolic HOI, by using the object-sensed tactile and kinematics data, without the information about pressure applied to the surface. By using the 10 hand-crafted features, for the classification of four emotions, we obtained an accuracy of 82.7%, and for two vitality forms, the accuracy is 84.85%. For emotion recognition, when only 7 seven tactile features are used, the aforementioned result stays the same while the corresponding F-score is only slightly worse (up to -0.19%). On the other hand, for vitality forms recognition, using only seven tactile features resulted in 7% lower performance, but when the best performing five features were extracted, four of them are tactile descriptors. The last results are particularly interesting. They show that even when reducing the number of sensors (to a tactile grid only), affect recognition is still possible. This makes this technology more affordable and applicable to various contexts. To the best of our knowledge, this is the first work that proposes computational approaches to deal with affect-related data of this type.

The main contributions of this paper are:

- We introduced a set of high-level easily interpretable features for tactile data to differentiate affective states.
- We show the feasibility of using machine learning methods to recognize emotions and vitality forms from the object-sensed tactile and kinematics data, even without the pressure data.
- Our results confirm that people touch the same object differently when performing basic actions such as rotating or grasping with different affective states.

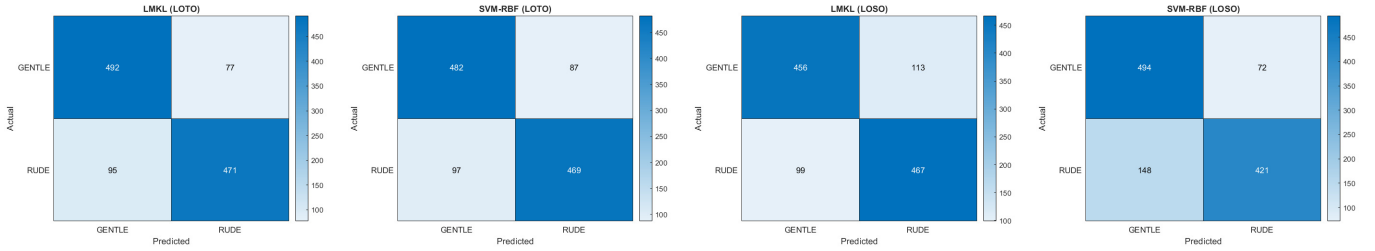

Fig. 3. Confusion matrices of LMKL and SVM-RBF for the classification of gentle and rude when 10 features are used and LOTO and LOSO are applied.

It is important to notice that these results were obtained for a semi-structured task enabling participants to interact in a personal and natural manner (i.e., with their "own style") without instructing them regarding how to perform the actions. Despite eventual interpersonal differences, the results are very promising. The second contribution highlights the role of touch modality during the communication of the vitality forms. We show, for the first time, that vitality forms can be automatically recognized from the way a person touches the surface of an object. This result complements the recent work on transmitting vitality forms with touch modality in human-human interactions [35]. Our approach can be used to classify different affective phenomena, and can be applied in daily-life tasks. Important to notice, that ecological validity is preserved in Study 2, in which the participants voluntarily perform actions communicating their attitudes.

Several future works are planned. Extending the dataset could allow us to apply other techniques of machine learning, aiming at finding better performing classifiers. Including additional data (e.g., linear acceleration), might potentially improve the classification results, and will be tested in the future. To evaluate the versatility of this approach, we will collect the data using objects of different physical properties (e.g., significantly smaller and lighter). We will also study whether it is possible to classify emotions, which are similar in terms of arousal and valence (e.g., anger, frustration and anxiety) as well as other vitality forms. We also plan to collect data for emotion classification in a more ecological setting.

Many existing systems and tools (including commercial ones) benefit from automatic recognition of human affective states. The other more futuristic applications are postulated by affective computing researchers. The application areas include well-being, personal development, and entertainment. We believe that our findings can contribute to the creation of new communication devices for people with reduced verbal communication [9], general purpose affect sensors for self-monitoring [10], [36], remote affective communication [37], video-games with affective feedback, intervention programs for neurodivergent persons who show reduced ability to perceive and communicate attitudes [38]. This technology can also be embedded in "smart" versions of several daily objects to sense the users' states, e.g., in the "smart-home" context.

ACKNOWLEDGMENTS

We are grateful to Prof. Giulio Sandini, Antonio Maviglia, Marcello Goccia, Diego Torrazza, Elio Massa and the others who ideated and developed the iCube, as well as to Serena Dominici and Linda Lovisolo. Alessandra Sciutti is supported by a Starting Grant from the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme, G.A. No 804388, wHiSPER.

REFERENCES

[1] J. Suchan and M. Bhatt, "Deep semantic abstractions of everyday human activities," in *ROBOT 2017: Third Iberian Robotics Conference*, A. Ollero, A. Sanfeliu, L. Montano, N. Lau, and C. Cardeira, Eds. Cham: Springer International Publishing, 2018, pp. 477–488.

[2] R. Wang, F. Quek, D. Tatar, K. S. Teh, and A. Cheok, "Keep in touch: Channel, expectation and experience," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI' 12. New York, NY, USA: ACM, 2012, pp. 139–148.

[3] M. Hertenstein, D. Keltner, B. App, B. Bulleit, and A. Jaskolka, "Touch communicates distinct emotions," *Emotion (Washington, D.C.)*, vol. 6, pp. 528–33, 09 2006.

[4] H. Lee Masson and H. Op de Beeck, "Socio-affective touch expression database," *PLOS ONE*, vol. 13, no. 1, pp. 1–21, 01 2018.

[5] R. Andreasson, B. Alenljung, E. Billing, and R. Lowe, "Affective touch in human–robot interaction: Conveying emotion to the nao robot," *Int. Journal of Social Robotics*, vol. 10, no. 4, pp. 473–491, Sep 2018.

[6] M. D. Cooney, S. Nishio, and H. Ishiguro, "Recognizing affection for a touch-based interaction with a humanoid robot," in *2012 IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, 2012, pp. 1420–1427.

[7] M. M. Jung, M. Poel, R. Poppe, and D. K. J. Heylen, "Automatic recognition of touch gestures in the corpus of social touch," *Journal on Multimodal User Interfaces*, vol. 11, no. 1, pp. 81–96, Mar 2017.

[8] Y. F. A. Gaus, T. Olugbade, A. Jan, R. Qin, J. Liu, F. Zhang, H. Meng, and N. Bianchi-Berthouze, "Social touch gesture recognition using random forest and boosting on distinct feature sets," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI '15. New York, NY, USA: ACM, 2015, pp. 399–406.

[9] D. Shapiro, Z. Zhan, P. Cottrell, and K. Isbister, "Translating affective touch into text," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19. New York, NY, USA: ACM, 2019, pp. LBW0175:1–LBW0175:6.

[10] F. Guribye, T. Gjøsæter, and C. Bjartli, "Designing for tangible affective interaction," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, ser. NordiCHI '16. New York, NY, USA: Association for Computing Machinery, 2016.

[11] K. Nakajima, Y. Itoh, Y. Hayashi, K. Ikeda, K. Fujita, and T. Onoye, "Emoballoon: A balloon-shaped interface recognizing social touch interactions," in *2013 IEEE Virtual Reality (VR)*, 2013, pp. 1–4.

[12] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?" *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 4, pp. 31:1–31:30, Dec. 2012.

[13] S. Ghosh, K. Hiware, N. Ganguly, B. Mitra, and P. De, "Emotion detection from touch interactions during text entry on smartphones," *Int. Journal of Human-Computer Studies*, vol. 130, pp. 47 – 57, 2019.

[14] Y. Sugiura, G. Kakehi, A. Withana, C. Lee, D. Sakamoto, M. Sugimoto, M. Inami, and T. Igarashi, "Detecting shape deformation of soft objects using directional photoreflectivity measurement," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 509–516.

[15] R. Niewiadomski and A. Sciutti, "Multimodal emotion recognition of hand-object interaction," in *26th International Conference on Intelligent User Interfaces*, ser. IUI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 351–355.
[16] K. R. Scherer, "What are emotions? and how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.
[17] A. Sciutti, F. Damonte, M. Alloisio, and G. Sandini, "Visuo-haptic exploration for multimodal memory," *Frontiers in Integrative Neuroscience*, vol. 13, p. 15, 2019.
[18] A. Sciutti and G. Sandini, "The role of object motion in visuo-haptic exploration during development," in *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, Aug 2019, pp. 123–128.
[19] H. G. Wallbott, "Bodily expression of emotion," *European Journal of Social Psychology*, vol. 28, no. 6, pp. 879–896, 1998.
[20] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Affective Computing and Intelligent Interaction*, A. Paiva, R. Prada, and R. Picard, Eds. Springer Berlin Heidelberg, 2007, pp. 71–82.
[21] S. C. Hauser, S. McIntyre, A. Israr, H. Olausson, and G. J. Gerling, "Uncovering human-to-human physical interactions that underlie emotional and affective touch communication," in *2019 IEEE World Haptics Conference (WHC)*, 2019, pp. 407–412.
[22] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161 – 1178, 1980.
[23] R. Hepach, D. Kliemann, S. Grüneisen, H. Heekeren, and I. Dziobek, "Conceptualizing emotions along the dimensions of valence, arousal, and communicative frequency – implications for social-cognitive tests and training tools," *Frontiers in Psychology*, vol. 2, p. 266, 2011.
[24] A. Ortony, G. Clore, and A. Collins, "The cognitive structure of emotion," vol. 18, 01 1988.
[25] D. A. Sauter, "The nonverbal communication of positive emotions: An emotion family approach," *Emotion Review*, vol. 9, no. 3, pp. 222–234, 2017, pMID: 28804510.
[26] M. Gönen and E. Alpaydin, "Localized multiple kernel learning," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: Association for Computing Machinery, 2008, pp. 352–359.
[27] M. Karg, K. Kühnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 1050–1061, 2010.
[28] R. Niewiadomski, M. Mancini, G. Varni, G. Volpe, and A. Camurri, "Automated laughter detection from full-body movements," *IEEE Trans. on Human-Machine Systems*, vol. 46, no. 1, pp. 113–123, Feb 2016.
[29] R. Niewiadomski, M. Mancini, S. Piana, P. Alborno, G. Volpe, and A. Camurri, "Low-intrusive recognition of expressive movement qualities," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ser. ICMI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 230–237.
[30] C. Beyan, V.-M. Katsageorgiou, and V. Murino, "Moving as a leader: Detecting emergent leadership in small groups using body pose," in *Proceedings of the 25th ACM International Conference on Multimedia*, ser. MM '17, 2017, pp. 1425–1433.
[31] C. Beyan, F. Capozzi, C. Becchio, and V. Murino, "Prediction of the leadership style of an emergent leader using audio and visual nonverbal features," *IEEE Trans. on Multimedia*, vol. 20, no. 2, pp. 441–456, 2018.
[32] C. Beyan, M. Shahid, and V. Murino, "Investigation of small group social interactions using deep visual activity-based nonverbal features," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM '18, 2018, pp. 311–319.
[33] D. N. Stern, *Forms of vitality exploring dynamic experience in psychology, arts, psychotherapy, and development*. Oxford University, 2010.
[34] G. Di Cesare, C. Di Dio, M. J. Rochat, C. Sinigaglia, N. Bruschweiler-Stern, D. N. Stern, and G. Rizzolatti, "The neural correlates of 'vitality form' recognition: an fMRI study," *Social Cognitive and Affective Neuroscience*, vol. 9, no. 7, pp. 951–960, 06 2013.
[35] G. Rizzolatti, A. D'Alessio, M. Marchi, and G. Cesare, "The neural bases of tactile vitality forms and their modulation by social context," *Scientific Reports*, vol. 11, 04 2021.
[36] F. Sarzotti, I. Lombardi, A. Rapp, A. Marcengo, and F. Cena, "Engaging users in self-reporting their data: A tangible interface for quantified self," in *Universal Access in Human-Computer Interaction*, M. Antona and C. Stephanidis, Eds., 2015, pp. 518–527.
[37] K. Woodward, E. Kanjo, S. Burton, and A. Oikonomou, "EmoEcho: A tangible interface to convey and communicate emotions," 2018.
[38] M. J. Rochat, V. Veroni, N. Bruschweiler-Stern, C. Pieraccini, F. Bonnet-Brilhault, C. Barthélémy, J. Malvy, C. Sinigaglia, D. N. Stern, and G. Rizzolatti, "Impaired vitality form recognition in autism," *Neuropsychologia*, vol. 51, no. 10, pp. 1918–1924, 2013.

Radoslaw Niewiadomski received the PhD degree in Computer Science from the University of Perugia (Italy). He is currently an Assistant Professor at the University of Trento. His research interests include emotion recognition, nonverbal behavior synthesis and multimodal interaction. He has been involved in several EU research projects, e.g., FP6 CALLAS, FP7 ILHAIRE and H2020 DANCE and co-authored over 75 peer-reviewed conference and journal papers.

Cigdem Beyan received her Ph.D. degree in Informatics from the University of Edinburgh, U.K., in 2015. She is currently an Assistant Professor at the University of Trento in the Department of Information Engineering and Computer Science. She has co-authored over 50 papers published in peer-reviewed journals and international conferences. Among her main research interest, there are human behavior understanding, social signal processing, and multimodal data analysis. She is a reviewer of several journals including various IEEE Transactions, and IEEE/ACM conferences. She is on the Editorial Board of ICES Journal of Marine Science covering the area of applications of computer vision and machine learning and was a Guest Co-Editor in Frontiers in Robotics and AI. She is a member of ELLIS.

Alessandra Sciutti is Tenure Track Researcher, head of the CONTACT (COgNiTive Architecture for Collaborative Technologies) unit of the Italian Institute of Technology (IIT). With a background on Bioengineering, she received her Ph.D. in Humanoid Technologies from the University of Genova in 2010. After two research periods in USA and Japan, in 2018 she has been awarded the ERC Starting Grant wHiSPER (www.whisperproject.eu), focused on the investigation of joint perception between humans and robots. She published more than 80 papers in international journals and conferences. She is currently Associate Editor for several journals, among which Cognitive Systems Research and the IEEE Transactions on Cognitive and Developmental Systems.