Pattern Recognition Letters 66 (2015) 31–40

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Recognizing affect in human touch of a robot ☆

Kerem Altun [a,*], Karon E. MacLean [b]

[a] Department of Mechanical Engineering, Istanbul Kemerburgaz University, Mahmutbey Dilmenler Cd. No: 26, Bağcılar, Istanbul, Turkey
[b] Department of Computer Science, The University of British Columbia, 2366 Main Mall, Vancouver, B.C. V6T 1Z4, Canada

ARTICLE INFO

Article history:
Available online 7 November 2014

Keywords:
Affective interfaces
Haptic
Human robot interaction
Affect recognition
Gesture recognition

ABSTRACT

A pet cat or dog's ability to respond to our emotional state opens an interaction channel with high visceral impact, which social robots may also be able to access. Touch is a key but understudied element; here, we explore its emotional content in the context of a furry robot pet. We asked participants to imagine feeling nine emotions located in a 2-D arousal-valence affect space, then to express them by touching a lap-sized robot prototype equipped with pressure sensors and accelerometer. We found overall correct classification (Random Forests) within the 2-D grid of 36% (all participants combined) and 48% (average of participants classified individually); chance 11%. Rates rose to 56% in the high arousal zone. To better understand classifier performance, we defined and analyzed new metrics that better indicate closeness of the gestural expressions. We also present a method to combine direct affect recognition with affect inferred from gesture recognition. This analysis provides a unique first insight into the nature and quality of affective touch, with implications as a design tool and for incorporating unintrusive affect sensing into deployed interactions.

1. Introduction

An interactive affective computing system requires automatic, realtime recognition of affect. In the past decade, extensive research in affect recognition has focused on vision, speech, physiological measurements, and their fusion [3,5,27]. However, the touch modality has not been widely considered, despite evidence that it is a potent means for emotion communication [11,12,14].

Recognition of different aspects of affect via touch will enable a substantially different approach to interactive applications, because of its potential for unintrusive use. Meanwhile, situated in an object rather than a space (as for vision), it can be built around interactions rather than a viewpoint from a fixed space. Touch-based affect sensing comprises a pipeline whose elements each impact ultimate recognition performance:

- the user's actual emotional state, its typical manifestation, and degree to which this is expressed in his/her touch;
- the object being touched, and the social and physical context of the interaction, both of which impact the expressiveness of the touch that is invited;
- data quality, namely the sensors used and their ability to detect expressively informative touches;
- recognition algorithm, delivering probabilities of a particular affective user state;
- metrics used to incorporate *a priori* known risk of misclassification.

Recently, good results have been obtained in touch-based affect recognition [20], which captures one interesting path through this large design space (Section 1.1).

In the present work, we explore the premise of human interactions with a haptically inviting and expressive social robot. In this context, interpretation of affective touch originating from the human, usually from a combination of endogenous origins and response to robot behavior, can potentially support closed-loop responsiveness in social human-robot interaction (HRI) settings, for example for therapeutic, caregiving and companionship applications with pet and humanoid robots. Our studies employ the Haptic Creature (Fig. 1), a robot pet that mimics a small animal sitting on a person's lap.

Approach: Our research, based on a robot pet platform that displays as well as senses touch stimuli, utilizes a dimensional representation of affect. The resulting structure should be helpful in robot sensing of human affect, by mitigating the impact of 'noisy' classification through a concept of 'near misses' and of natural, i.e. higher-likelihood transitional paths; and in robot rendering of its own affect display, via a topological map of coherent dynamic transitions. In a previous study, participants performed a series of specified touch gestures on the robot as they imagined feeling different emotions, and touch data were recorded using first-generation touch sensors. An analysis of the visual and video observation of these gestures has been published [26]. However, these methods come at a high cost of labor,

☆ This paper has been recommended for acceptance by Friedhelm Schwenker.

* Corresponding author. Tel.: +90 212 604 0100; fax: +90 212 445 9255.
E-mail address: kerem.altun@kemerburgaz.edu.tr (K. Altun).

http://dx.doi.org/10.1016/j.patrec.2014.10.016
0167-8655/

Fig. 1. The Haptic Creature, in experiment context.

processing delay and the intrusiveness and privacy issues of video recording.

Here, we use pattern recognition on this study's unpublished touch sensor data to assess feasibility and best methods for classifying the gesture-giver's affective state, and to generate requirements for next-generation touch sensing. We consider two schemes: (1) recognizing affect directly from touch data, and (2) identifying touch gestures *and* inferring affect information from gesture, then combining these two information sources.

1.1. Related work

1.1.1. Affective content in touch gestures

The first item in our pipeline indicates the central assumption that the user's touch is emotionally informative. Evidence that it can be comes from multiple sources. Touch has long been observed in real situations by social scientists to encode social meaning, albeit modified by factors such as relationship, hierarchy and context [13]; and studies in nursing chronicle its use for healing, *e.g.*, [1]. In controlled environments, individuals asked to imagine feeling varying emotions are observed to make different touch gestures [26], and those asked to *express* different emotions to a human or mannequin arm make gestures and touch patterns that can be distinguished and which correlate to the instructed emotion [20].

It remains to validate identifiable information content in touches made under authentic emotional circumstances. This effort will be assisted in future by a system able to capture touch data automatically along with other verifying context (e.g. voice prosody, caregivers' reports). The present work is a step toward that end.

1.1.2. Affective classification of touch gestures

Affect representations in use today generally take one of two forms. A *categorical* approach models affective states as independent classes; a *dimensional* approach organizes them with a systematic relation [10]. Of the latter, the most well-known is Russell's circumplex model of affect [15], which locates emotions on dimensions of valence (x-axis) and arousal (y-axis). Following the experiment design under which our data were collected, fully motivated in [25], we use a modification of the circumplex model known as the *affect grid* ([16], Fig. 2). Further discussion and review of continuous emotion models can be found in [10], and in [9] in the context of affect recognition, and we refer the reader to Silvera-Tawil et al. [20] for a comprehensive review on tactile HRI.

e_1 distressed	e_2 aroused	e_3 excited
e_4 miserable	e_5 neutral	e_6 pleased
e_7 depressed	e_8 sleepy	e_9 relaxed

Fig. 2. Emotion labels (from [16]) and their locations in the affect grid, as adapted by Yohanan and MacLean [26].

Past efforts have typically employed two steps to recognize affect out of touch, first identifying the gestures and then attaching an affective meaning to those gestures. For example, the robot seal PARO [19] can sense if it is being stroked *vs.* hit through its touch sensors. Huggable's more sophisticated algorithm and structure distinguishes nine touch gestures [21]. Probo defines three gestures at a higher level—a hug, scratch, or hurt [17]; AIBO detects only touch location of touch [8]. All these robots employ some kind of gesture recognition as part of affective communication.

Most relevantly and recently, Silvera-Tawil et al. [20] asked subjects to display specified emotion messages to a real human arm, and to a mechanical proxy with touch sensors. Six emotion states were recognized with accuracy comparable to humans (47%–52%) representing to our knowledge a first instance of machine recognition of affective content in touch.

Our study, while sharing a general aim, differs throughout the recognition "pipeline". We employed a human-pet paradigm: touching a nice-feeling animal proxy is socially very different from touching a human or its proxy. Rather than *communicate* a specific emotion, our subjects were instructed to *imagine* different emotional states, and to interact as if with a real pet companion; it is unclear at this stage whether this makes the classification problem harder (display is not explicit) or easier (more natural circumstance makes differences more clear). Sensing was accomplished with discrete transducers rather than a continuous sensed sheet, producing discontinuous data which likely impacted ultimate performance. Finally, we implemented metrics based on a dimensional representation of emotion, allowing us to apply distances to the topological connections between actual and detected affective states. To our knowledge, recognizing and classifying affective touch in such a setting is new.

1.2. Research questions and contributions

This study aims to answer, in the context described above:

- How far can we push the limits of inexpensive sensors by applying advanced machine learning methods to extract affect information from touch?
- Is a dimensional (*e.g.*, 2-D arousal/valence space) representation informative for affective touch?
- What are good performance metrics for dimensional affect recognition?
- Which emotion states of those examined here tend to be confused by our algorithm and sensor?
- Is inferring affect directly from gestures more or less effective than recognizing gestures then attaching emotional meaning to them, for the conditions studied here?

Our original contributions are the following:

- Insights into the feasibility of classifying affect directly from touch data, in the realistic case where the touch-giver is not explicitly trying to communicate it.

Fig. 3. Robot surface showing placements of 56 force sensing resistors (FSR) on robot body. For any given gesture, data from the three most informative FSRs were used.

- Metrics defined in affect space to evaluate and compare recognition performance for emotional states.
- A simple method to combine gesture recognition information with affect recognition.

2. Methodology

2.1. Experimental setup

The Haptic Creature's perception comes from a frame-mounted accelerometer and a network of 56 surface-mounted force sensing resistors (FSR; Fig. 3). Its sensing capabilities are restricted by the limited coverage of the FSRs, the best technology available at time of construction (2008). Furthermore, FSRs are sensitive to forces normal to the surface, whereas many touch gestures include tangential (shear) forces. The accelerometer can detect but not localize both robot motion and the small vibrations associated with a touch.

Here we analyze touch and accelerometer data collected experimentally, to get the most out of these inexpensive sensors using a sophisticated machine learning algorithm. This in turn provides insight into requirements for more adequate touch sensing. Details on the mechatronic design of the Haptic Creature, including the specifications for the touch sensors can be found in [24].

2.2. Experimental procedure

We analyzed previously unreported touch data collected from 31 participants as part of the study fully reported in [26]; here, we briefly summarize the procedure. Consistently with previous Creature studies, we used nine emotion labels representing a 3×3 affect grid (Fig. 2), where the horizontal axis is valence (negative to positive), and the vertical axis is arousal (low to high), with a *neutral* state at the middle. The specific emotion labels used are *distressed, aroused, excited, miserable, neutral, pleased, depressed, sleepy,* and *relaxed*, and were ultimately derived from Russell's circumplex labels [16].

The experiments were conducted in a room with a desktop computer and a video camera observing the participant. The Haptic Creature was placed on the participant's lap initially; but she was free to adjust its position throughout the study. The experiment facilitator was not present in the room with the participant; the instructions were displayed and the data were collected by the computer. At the start of the experiment, the participant was presented with brief information about the Haptic Creature, such as how it moves and how it senses the surroundings through touch. She was instructed to *"imagine the Haptic Creature to be [her] pet, one with which [she has] a close and comfortable relationship."* She was further instructed that the robot will be passive and will not react in any way. Then, the participant was presented with emotion labels in random order, and for each emotion label, she was asked to rate the likelihood of performing 30 different gestures selected from Yohanan's gesture library [26]. The rating scale was defined in the range: 1 (Very Unlikely) to 5 (Very Likely). Then, for gestures reported with a likelihood of 4 or 5, the participant was asked to perform that gesture on the Haptic Creature, imagining that she was feeling that emotion. Each gesture was captured through video, and force and acceleration data were recorded by the robot's FSRs and accelerometer. Video analysis results are reported in [26]. Here, we analyze touch and accelerometer data alone, and use survey results to estimate prior probabilities in combining the gesture recognition information with affect recognition (Section 2.4). We use 26 gestures; the four least likely gestures were discarded because of insufficient performance data. More details about the experimental procedure can be found in [23].

2.3. Data processing and feature extraction

Instead of working with raw time series data (average duration 8.8 s, $\sigma = 4.8$ s), we calculated several features which have proven effective for touch gesture recognition [7]. These comprised the time series' mean, median, variance, minimum, maximum, and total variation. In addition, we calculate the series' Fourier transform and use its peak and the corresponding frequency as features. Thus, we calculate eight features per signal. No normalization is applied to the signals or the features.

We found that for most of the data only a few FSRs activated; this may have been due to both actual participant activity, and limited force sensing capabilities. For each gesture, we therefore calculated the variances of the FSR signals and discarded all but the three with the largest variance for that gesture; this produced $(3 \times 8 = 24)$ FSR features. Position of the active sensors was not used in the feature, meaning our features were location-independent. For example, if two people perform the exact same gesture but at different locations on the robot, the features that we calculate will exactly be the same.

While this is essential for gesture recognition, one might argue that some information is lost for emotion recognition, since emotion and touch location are correlated. This might be the case in human-human(oid) interaction, but in our case of human-pet interaction, associating touch locations with emotional states has not been reported before, to our knowledge. Our case is inherently different; in human-human touch interaction, the primary aim usually is to convey some emotional information to the recipient. However, in human-pet interaction, humans rather interact with the pet in a more relaxed manner and do not try to convey an emotion. This arguably could render touch location in human-pet interactions less important. In our experiments, we observed that introducing location feature did not improve the results so we disregarded the location data.

The three-axis accelerometer provided another three channels from which we computed the same features as for the FSRs, or $(3 \times 8 = 24)$. This resulted in 48 features in total.

2.4. Classification

We use Random Forests (RF) as the classifier in this study [2], a method based on combining classification results from multiple decision trees. It gives the best results among the other classifiers that we considered (Fisher linear discriminant, naive Bayes, *k*-nearest neighbor, decision tree, multilayer perceptron). Furthermore, this method has proven most effective among many alternatives considered for gesture recognition [6] and gait classification [18].

We use the RF classifier to recognize affect from sensor data, *i.e.*, to generate a probability distribution (PD) that the sensor data belongs to a given emotion class e_i. In the following, we denote this PD obtained for direct affect recognition by $P_{E(\mathrm{RF})}(e_i|x)$, where x represents the

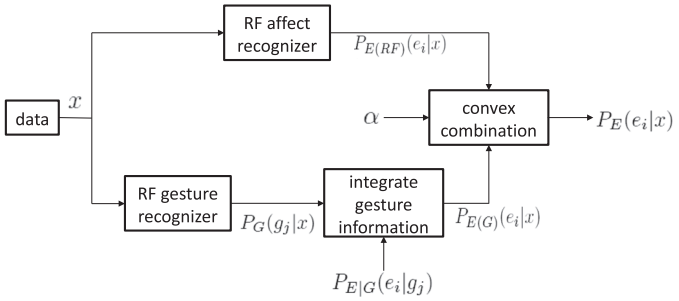

Fig. 4. Combining gesture recognition information with affect recognition.

data. After this PD is obtained, x can be assigned the class with the highest probability.

We now consider the relation between gesture and emotional states, and how to exploit it for affect recognition. Silvera-Tawil et al. [20] relate their touch gestures ("touch modalities") to the messages that subjects have been instructed to convey. In our study, the gestures are performed freely by participants in an imagined emotional state. Here, we introduce a way to integrate the gesture information into affect recognition and analyze the contribution of gestures to the perceived emotional state.

Assume (Fig. 4) that a prior PD over emotional states for each given gesture g_j is available. Let $P_{E|G}(e_i|g_j)$ denote the probability of feeling emotion e_i given that gesture g_j is performed. This probability is not related to the collected gesture/affect sensor data, but is simply prior information that we assume we know before analyzing (or even collecting) the data.

If the prior probabilities are known, gesture recognition can provide information about the emotional state. Therefore, we also run the RF algorithm for gesture recognition, without regard to emotional state ("RF gesture recognizer" block in Fig. 4). Let $P_G(g_j|x)$ be the output of the RF classifier indicating the probability that data x belongs to gesture class g_j, obtained by normalizing the number of votes for each class. Then, affect can be inferred from this information:

$$P_{E(G)}(e_i|x) = \sum_{j=1}^{26} P_{E|G}(e_i|g_j)P_G(g_j|x) \tag{1}$$

where $P_{E(G)}(e_i|x)$ is the PD of the emotional state inferred from gesture information ("integrate gesture information" block in Fig. 4). We denote this distribution by $P_{E(G)}(e_i|x)$ to discern it from the RF distribution used for affect recognition directly (i.e., without going through gestures; "RF affect recognizer" block in Fig. 4), denoted by $P_{E(\mathrm{RF})}(e_i|x)$ above. These two distributions can be combined to give the final estimate $P_E(e_i|x)$. Many methods can be applied for combination [4] but the simplest is the convex combination:

$$P_E(e_i|x) = \alpha P_{E(\mathrm{RF})}(e_i|x) + (1-\alpha)P_{E(G)}(e_i|x) \tag{2}$$

where $0 \le \alpha \le 1$ is a specified constant. It can be selected according to how reliable each estimate is.

2.5. Classifier evaluation based on a choice of loss functions

In this section, we describe the methods we use to evaluate the performance of the classifier and make a performance comparison for different emotional states. We first introduce the notion of a loss function, then show how different loss functions can be used to apply distances to the topological space defined by the dimensional emotion model, based on one's purpose or *a priori* knowledge. Finally we illustrate six representatives with which we will analyze this dataset.

As opposed to saying that a classification is either correct or incorrect, the loss function assigns a degree to an incorrect classification. For a correct classification, the loss is zero; but for an incorrect one, it is non-negative. This number assigns a cost to that incorrect classification.

For example (see Fig. 2), suppose that a classifier incorrectly classifies *depressed* states as *miserable* and also incorrectly classifies *relaxed* states as *distressed*. The classifier's performance is thus bad for both *depressed* and *relaxed* states. However, it should not be equally bad, because the distance between *depressed* and *miserable* is much less than between *relaxed* and *distressed*. Therefore, by defining loss functions, we can compare performances of classifying different emotional states not by the correct classification percentage, but considering the degree of misclassification. We define several loss functions in the following and measure the performance of the random forest classifier for different emotions.

Let $\mathcal{E} = \{e_1, \ldots, e_9\}$ be the set of emotions. Fig. 2 shows the emotions and their positions in the affect space. Let the number of samples in class i be denoted as N_i; the loss incurred by deciding e_j when the actual class is e_i as $L(e_j; e_i)$; and the probability of deciding e_j when the actual class is e_i as $P(e_j; e_i)$. $P(e_j; e_i)$ can then be estimated by dividing the number of samples actually belonging to class e_i that are classified as e_j by the total number of samples belonging to class e_i, after testing a classifier by cross validation. So,

$$P(e_j; e_i) = \frac{\text{number of samples in } e_i \text{ classified as } e_j}{N_i} \tag{3}$$

Expected loss for emotion class i and loss function L is then

$$\mathcal{L}_i(L) = \sum_{j=1}^{9} L(e_j; e_i)P(e_j; e_i) \quad i = 1, \ldots, 9 \tag{4}$$

2.5.1. A discrete (binary) loss function

A trivial choice for the loss function is the discrete (0-1) loss function L_D that simply declares the loss zero for correct classification, otherwise one. This obviously does not account for the degree of misclassification. Therefore, we define other loss functions based on different distances in affect space.

2.5.2. A distance-based loss function

If we represent emotions as points in two dimensions, we can define the *distance* between the emotional states e_i and e_j in the affect space. Now, the expected loss in Equation (4) accounts for degree of misclassification. Most simply, we can use the Euclidean distance, given by

$$L_E(e_j; e_i) = \kappa\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{5}$$

where κ is a scaling factor and (x_i, y_i) are the valence and arousal coordinates of an emotional state, respectively. The subscript emphasizes the Euclidean distance, and κ is chosen such that the loss for the worst misclassification (diagonal of the square) is 1. Fig. 5 shows distances for *distressed*, *aroused*, and *neutral* states, which together capture the distances for all nine emotional states, available upon rotation or reflection.

For compatibility, we can also treat the discrete loss function L_D as a distance. All nonzero distances then equal 1.

2.5.3. More tolerant loss functions

It is possible to define more "tolerant" distances in the affect space and use them as loss functions to evaluate classification performance,

0	0.35	0.71
0.35	0.5	0.79
0.71	0.79	1

(a)

0.35	0	0.35
0.5	0.35	0.5
0.79	0.71	0.79

(b)

0.5	0.35	0.5
0.35	0	0.35
0.5	0.35	0.5

(c)

Fig. 5. Euclidean distances for (a) e_1, (b) e_2, and (c) e_5 in the affect grid.

0	0.13	0.5
0.13	0.25	0.63
0.5	0.63	1

(a)

0.13	0	0.13
0.25	0.13	0.25
0.63	0.5	0.63

(b)

0.25	0.13	0.25
0.13	0	0.13
0.25	0.13	0.25

(c)

Fig. 6. Squared Euclidean distances for (a) e_1, (b) e_2, (c) e_5 on affect grid.

and use them to account for the inherent fuzziness in affect. For example, a loss function can be defined to tolerate the misclassification if it is a "near miss," *i.e.*, the actual emotion and the classified emotion are neighbors in the affect space. This has been done by Wöllmer et al. [22] in the context of affect recognition from speech.

An extreme example would assign zero loss if the actual emotion and the classified emotion are neighbors in the affect space. However, this is likely too tolerant. Instead, we would like something in between, with a small loss for adjacent emotional states, but a large one for distant states. The square of the Euclidean distance would satisfy these conditions:

$$L_{E^2}(e_j; e_i) = \lambda[(x_i - x_j)^2 + (y_i - y_j)^2] \tag{6}$$

where λ is a scale factor; values are given in Fig. 6. For consistency, λ is chosen such that the maximum loss value is 1.

2.5.4. Relation to emotion models and another loss function

It is possible to link the loss functions defined above to emotion models in the literature. For example, the discrete metric L_D considers each emotion as a separate entity, i.e., it does not take into account the distance in the emotion space. Therefore, it can be considered as a representation of Ekman's view of emotion in the affect recognition domain. Similarly, the metric L_E places each emotion on Russell's affect grid, so it is a representation of that approach. A more popular emotion model is Russell's circumplex. Here, the emotion e_5 (*neutral*) would be at the origin and other emotions would roughly form a circle centered at the origin. We assume that the remaining eight emotions are equally spaced around the circle. Assuming a circle with diameter 1, the loss values are shown in Fig. 7. We denote that loss function as L_C.

2.5.5. Loss functions treating arousal and valence separately

In order to analyze the classification performance separately in arousal and valence dimensions, we define two other loss functions. L_V focuses on the valence dimension. If the true and classified emotions are in the same valence zone, the loss is zero; if the valence is opposite, the loss value is 1; otherwise it is defined as 0.5. L_A similarly focuses on the arousal dimension. Note that for L_V, the loss along the arousal axis is zero, and for L_A, the loss along the valence axis is zero. Therefore, the higher L_V is, the worse the performance of a classifier in valence dimension; the higher L_A is, the worse the performance of a classifier in arousal dimension. The loss values for L_V and L_A are given in Figs. 8 and 9, respectively.

We note here that all metrics defined above are normalized such that the maximum possible loss value is 1. This enables fair comparison between these metrics. For example, compare L_{E^2} to the loss function defined by the Euclidean metric L_E. If two emotions e_i and e_j are adjacent, $L_E(e_j; e_i) = 1/2\sqrt{2}$ but $L_{E^2}(e_j; e_i) = 1/8$. So L_{E^2} tolerates a near miss more than the Euclidean metric. If e_i and e_j have the same valence but opposite arousal, $L_E(e_j; e_i) = 1/\sqrt{2}$ and $L_{E^2}(e_j; e_i) = 1/2$.

0	0.38	0.71
0.38	0.5	0.92
0.71	0.92	1

(a)

0.38	0	0.38
0.71	0.5	0.71
0.92	1	0.92

(b)

0.5	0.5	0.5
0.5	0	0.5
0.5	0.5	0.5

(c)

Fig. 7. Circumplex model distances for (a) e_1, (b) e_2, and (c) e_5.

0	0.5	1
0	0.5	1
0	0.5	1

(a)

0.5	0	0.5
0.5	0	0.5
0.5	0	0.5

(b)

0.5	0	0.5
0.5	0	0.5
0.5	0	0.5

(c)

Fig. 8. L_V distances for (a) e_1, (b) e_2, (c) e_5.

0	0	0
0.5	0.5	0.5
1	1	1

(a)

0	0	0
0.5	0.5	0.5
1	1	1

(b)

0.5	0.5	0.5
0	0	0
0.5	0.5	0.5

(c)

Fig. 9. L_A distances for (a) e_1, (b) e_2, (c) e_5.

Clearly, it is possible to define many other loss functions. However, these model the most frequently used dimensional approaches to affect in psychology, and capture the nature of the misclassifications. Different loss functions might be appropriate for other applications.

Here we point out that the loss function L_{E^2} is not a mathematical distance as it violates the triangle inequality. However, we observe that this allows it to model the case where the affect space is not a proper metric space. For example, according to this loss function e_1 is close to e_2, e_2 is close to e_3, but e_1 and e_3 are far apart.

Among metrics that are not proper distances, we highlight the non-symmetric loss functions. When $L(e_j; e_i) \neq L(e_i; e_j)$, the loss in classifying e_i as e_j will not be equal to the loss in classifying e_j as e_i. This can have interesting utilities. For example, in a therapeutic study, it might be most crucial to classify negative emotions correctly, with the highest cost associated with an incorrect positive classification. However, non-symmetric loss functions are beyond our scope.

3. Analysis and discussion

3.1. Direct affect recognition

We first report results with direct affect recognition, *i.e.*, with touch gesture labels disregarded. A test pattern x is then assigned to the highest-probability emotion class. More precisely,

$$\text{label}(x) = \arg\max_i P_{E(\text{RF})}(e_i|x). \tag{7}$$

This is the output of Fig. 4's "RF affect recognizer" block.

3.1.1. Results with all participants' data

Affect recognition results with 10-fold cross validation are presented as a confusion matrix in Table 1, where rows represent the true and columns the estimated class. Overall correct recognition is 35.9%; chance is 11%.

This recognition rate does not take the closeness of emotional states into account. For example, in row 3, 164 of the e_3 (*excited*) samples are correctly classified. The next highest number in the row, 54, corresponds to column 6. This means that 54 of the e_3 (*excited*) samples are incorrectly classified as e_6 (*pleased*), which is in fact right next to e_3 in the affect space. It is hard to observe these "near misses" in the confusion matrix.

To better illustrate the real situation, we also present the same results in graphical form. Fig. 10(a) shows our visualization for the confusion matrix in Table 1. Each shaded 9-cell block in the figure corresponds to a row in the confusion matrix, according to the labels defined in Fig. 2. For example, the position "medium arousal, positive valence" in Fig. 10(a) corresponds to row 6 (e_6) in the confusion matrix. Within each block, we represent the classification results for that row, again according to Fig. 2 labels as follows. First, each row e_i is

Table 1
The confusion matrix for affect recognition with all participants' data combined. Note that e_{1-9} reside in a 2-dimensional (3 × 3) space, making this "flattened" view hard to interpret. See Fig. 10 for a more intuitive 3-D visualization.

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9
e_1	51	22	44	5	13	36	15	15	22
e_2	9	118	58	6	21	46	10	12	28
e_3	22	39	164	6	16	54	9	8	19
e_4	19	21	35	36	18	26	14	12	29
e_5	6	20	27	14	92	42	13	21	40
e_6	16	46	80	5	24	137	18	5	22
e_7	9	23	26	15	10	35	72	17	30
e_8	15	21	17	5	26	25	14	76	34
e_9	14	15	25	11	26	31	16	18	142

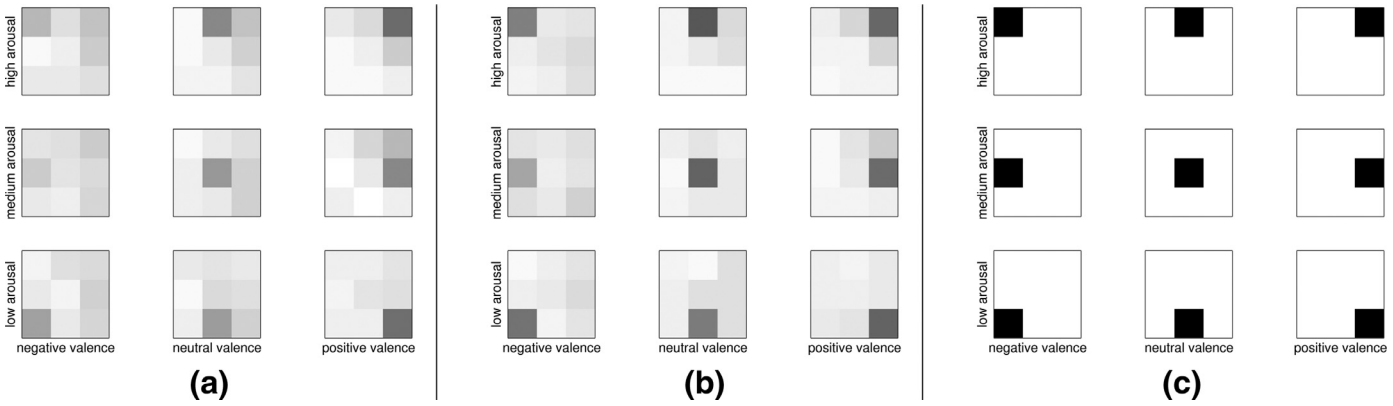

Fig. 10. Visual quasi-3-D representation of affect recognition results (probabilities). For example, each upper-left block shows the 3x3 distribution of recognition results when the high-arousal, negative valence emotion word (*distressed*) was imagined while the participant gestured, and so on. (a) All participants' data combined (values listed in Table 1); (b) average of individual participants; (c) hypothetical perfect classification.

divided by the sum of the numbers in that row, to estimate the classification probabilities $P(e_j; e_i)$ (probability of deciding the class is e_j when the actual class is e_i) defined in Equation (3). Then, these probabilities are arranged in a 3 × 3 structure consistent with the labeling in Fig. 2 and represented by cell, with shade darkness proportional to probability. Fig. 10(c) illustrates a hypothetical perfect classification.

This representation is particularly useful for visualizing the near misses in the confusion matrix. Two key observations now emerge: first, in Fig. 10(a), most of the incorrectly classified e_3 (*excited*) state samples are classified as the neighboring states, namely e_2 (*aroused*) and e_6 (*pleased*), *i.e.* they are near misses. Second, classification results are weakest for e_1 (*distressed*) and e_4 (*miserable*) - upper left and middle left, respectively: most of the samples are incorrectly classified (relatively uniform grey across block), and moreover, many incorrect classifications are nonadjacent. Interestingly, this is consistent with Yohanan's observations that it is hard to render emotion in negative valence and medium-high arousal zone [23].

To quantify classification performance of different emotional states, we can use the distance metrics defined in the previous section to systematically compare performance for each affect-grid region among each other and to baseline values, which differ by metric and location on grid. In Table 2, for each of our distance metrics we present expected loss values and results for individual and baseline performance. Two baseline classifiers are considered for comparison. *Class frequency*: assigns a class to a sample with a probability proportional to the frequency of that class in the data set. *Chance*: classifies a sample randomly; *i.e.*, each emotion has a 1/9 chance (11.1%) of assignment to a given sample. These two classifiers represent a baseline that any classifier is expected to beat, because they use no analysis. *Worst case*: assigns a sample to the class for which the loss would be maximized; *e.g.*, if a sample's correct class is e_9 (*relaxed*), it is assigned to e_1 (*distressed*). This is not a proper classifier because it uses class information, *i.e.*, it already "knows" the true class of a sample and assigns it to the worst class possible.

In viewing Table 2, keep in mind first, that the random guess and worst case values vary by location in the affect grid because the number of neighbors differ; and secondly, the L_D (distance-based loss function) values are in fact error rates (fraction of misclassified samples) for each class and can be interpreted as the probability of misclassification.

Clearly, a lower loss value means better performance. From Table 2 results, we can conclude that performance in the positive valence zone is generally better. For example, for the e_3 (*excited*) state, $(1 - 0.51 =)$49% of samples are classified correctly (second column, loss function L_D). Interestingly, for e_1 (*distressed*), e_4 (*miserable*), and e_7 (*depressed*), L_V is significantly higher than L_A. For negative valence emotions, near misses in arousal are much more likely than near misses in valence; *i.e.*, misclassifications of arousal are more likely to be near misses than those of valence. Similarly, for e_8 (*sleepy*) and e_9 (*relaxed*), L_A values are higher than L_V values. These emotions have more near misses in valence than in arousal.

L_D values are the highest among all other loss functions (even the worst case classifier) because near misses are not reflected in L_D. To compare near-miss rate for different emotions, we calculate percentage decrease in loss value for L_E, L_{E^2}, L_C, L_V, and L_A as compared to L_D for the RF classifier, and for class frequency (CF) as a baseline. Differences between the RF and CF classifier percentage decreases are shown in Fig. 11(a).

This chart displays a measure of the distribution of just the incorrect classifications. If most incorrect classifications are near misses and their number is larger than that of the class frequency classifier, the bar is higher. This view shows the highest number of near misses

Table 2
Expected loss values, for (column 2 of each matrix) Random Forests trained on all participants' data and (column 3) on individual participants' data then averaged. Baseline data are (column 4) the class frequency classifier, (column 5) random guess classifier, and (column 6) worst case classifier. Each matrix shows data computed for a different loss function: (a) L_D, (b) L_E, (c) L_{E^2}, (d) L_C, (e) L_V, (f) L_A.

Emotion	All	Individual	Class frequency	Chance	Worst case	Emotion	All	Individual	Class frequency	Chance	Worst case
(a) Discrete loss function (L_D)						(d) Circumplex distance loss function (L_C)					
e_1	0.77	0.59	0.91	0.89	1.00	e_1	0.57	0.43	0.64	0.61	1.00
e_2	0.62	0.44	0.88	0.89	1.00	e_2	0.39	0.26	0.60	0.61	1.00
e_3	0.51	0.49	0.86	0.89	1.00	e_3	0.28	0.27	0.57	0.61	1.00
e_4	0.83	0.71	0.92	0.89	1.00	e_4	0.62	0.51	0.66	0.61	1.00
e_5	0.67	0.47	0.89	0.89	1.00	e_5	0.33	0.24	0.44	0.44	0.50
e_6	0.61	0.50	0.86	0.89	1.00	e_6	0.35	0.30	0.57	0.61	1.00
e_7	0.70	0.53	0.90	0.89	1.00	e_7	0.52	0.39	0.66	0.61	1.00
e_8	0.67	0.57	0.91	0.89	1.00	e_8	0.44	0.37	0.63	0.61	1.00
e_9	0.52	0.47	0.88	0.89	1.00	e_9	0.33	0.31	0.59	0.61	1.00
average:	0.66	0.53	0.89	0.89	1.00	average:	0.43	0.34	0.60	0.60	0.94
(b) Euclidean distance loss function (L_E)						(e) Valence loss function (L_V)					
e_1	0.54	0.40	0.60	0.58	1.00	e_1	0.57	0.40	0.56	0.50	1.00
e_2	0.31	0.20	0.47	0.48	0.79	e_2	0.25	0.17	0.34	0.33	0.50
e_3	0.27	0.26	0.54	0.58	1.00	e_3	0.20	0.22	0.44	0.50	1.00
e_4	0.49	0.41	0.52	0.48	0.79	e_4	0.55	0.43	0.56	0.50	1.00
e_5	0.28	0.20	0.38	0.38	0.50	e_5	0.26	0.16	0.34	0.33	0.50
e_6	0.28	0.24	0.45	0.48	0.79	e_6	0.22	0.19	0.44	0.50	1.00
e_7	0.49	0.37	0.62	0.58	1.00	e_7	0.49	0.37	0.56	0.50	1.00
e_8	0.35	0.30	0.50	0.48	0.79	e_8	0.24	0.22	0.34	0.33	0.50
e_9	0.32	0.30	0.56	0.58	1.00	e_9	0.24	0.26	0.44	0.50	1.00
average:	0.37	0.30	0.51	0.51	0.85	average:	0.34	0.27	0.45	0.44	0.83
(c) Squared Euclidean distance loss function (L_{E^2})						(f) Arousal loss function (L_A)					
e_1	0.40	0.30	0.44	0.42	1.00	e_1	0.35	0.32	0.48	0.50	1.00
e_2	0.17	0.10	0.28	0.29	0.63	e_2	0.28	0.17	0.48	0.50	1.00
e_3	0.16	0.16	0.37	0.42	1.00	e_3	0.22	0.20	0.48	0.50	1.00
e_4	0.32	0.26	0.32	0.29	0.63	e_4	0.31	0.29	0.33	0.33	0.50
e_5	0.12	0.09	0.17	0.17	0.25	e_5	0.23	0.19	0.33	0.33	0.50
e_6	0.15	0.12	0.26	0.29	0.63	e_6	0.26	0.21	0.33	0.33	0.50
e_7	0.37	0.28	0.46	0.42	1.00	e_7	0.37	0.29	0.52	0.50	1.00
e_8	0.20	0.17	0.30	0.29	0.63	e_8	0.35	0.30	0.52	0.50	1.00
e_9	0.21	0.21	0.39	0.42	1.00	e_9	0.30	0.24	0.52	0.50	1.00
average:	0.24	0.19	0.33	0.33	0.75	average:	0.30	0.24	0.44	0.44	0.83

for neutral to positive valence emotions, especially for e_3 (*excited*) and e_6 (*pleased*). For negative valence emotions ($e_{1,4,7}$), the negative bar value indicates that a class frequency classifier is expected to have more near misses. Taken together, this means that the incorrect classifications of the RF classifier are distributed away from the actual emotion for negative valence emotions, but they are distributed close to the actual emotion for neutral and positive valence emotions.

Among different loss functions, L_E, L_{E^2}, and L_C values are close, which means that the way that the distance is defined is not very significant. However, interesting conclusions can be drawn by observing L_V and L_A bars. The L_A bar is highest for high arousal emotions ($e_{1,2,3}$). Since L_A is zero in the valence dimension, this bar measures the performance in the arousal when valence is disregarded. For this case, the classification performance is the best for the high arousal emotions $e_{1,2,3}$. Similarly, L_V measures the performance in the valence dimension when arousal is disregarded. For this case, performance is the best for positive valence emotions ($e_{3,6,9}$). Therefore, we can conclude that near misses in arousal are more likely to occur for high arousal emotions; near misses in valence are more likely to occur for positive valence emotions (upper side and right side of the square-shaped affect space).

3.1.2. Results for individual participants

Affective communication between a robotic pet and its user is usually idiosyncratic; individual affect recognition results are thus also of interest. We run the RF classifier for each participant, normalize each resulting confusion matrix (by dividing each row by its sum) to get probabilities, then average them over all participants. A visual representation is given in Fig. 10(b); expected loss values appear in the third column of Table 2. The average correct recognition rate for this case is 47.8%.

As expected, loss values are usually lower for individual participants. A similar calculation for the percentage decrease in loss can be calculated for this case as well (Fig. 11(b)). Performance for negative valence emotions is better than when trained on all participants, but the class frequency classifier still slightly outperforms Random Forests. Results are qualitatively similar to the case with all participants: near misses are more likely in arousal for e_2 and e_3, and in valence for e_3 and e_6. The exception is that near misses in arousal are also likely for e_9.

3.2. Combining gesture information with affect recognition

We also applied the RF method for gesture recognition and combined the results with direct affect recognition. In this case, affect is classified according to the combined PD:

$$\text{label}(x) = \arg\max_i P_E(e_i|x). \tag{8}$$

For gesture recognition, we were able to assess classifier performance since participants were instructed not only to imagine an emotion, but also to perform a gesture and this was thus known *a priori*. We disregarded instructed emotional state and trained the classifier to distinguish directly between all 26 gestures. This gives the probability distribution $P_G(g_j|x)$ for a test sample x. Assigning the gesture class with the highest probability to x, the confusion matrix for 10-fold cross validation is presented in Table 3. Overall correct classification is 32.8%, an order of magnitude larger than chance ($1/26 = 3.9\%$). Further, most confusions occur within similar gestures.

We estimated prior probabilities $P_{E|G}(e_i|g_j)$ in Equation (1) from the survey collected before gesture acquisition, which asked

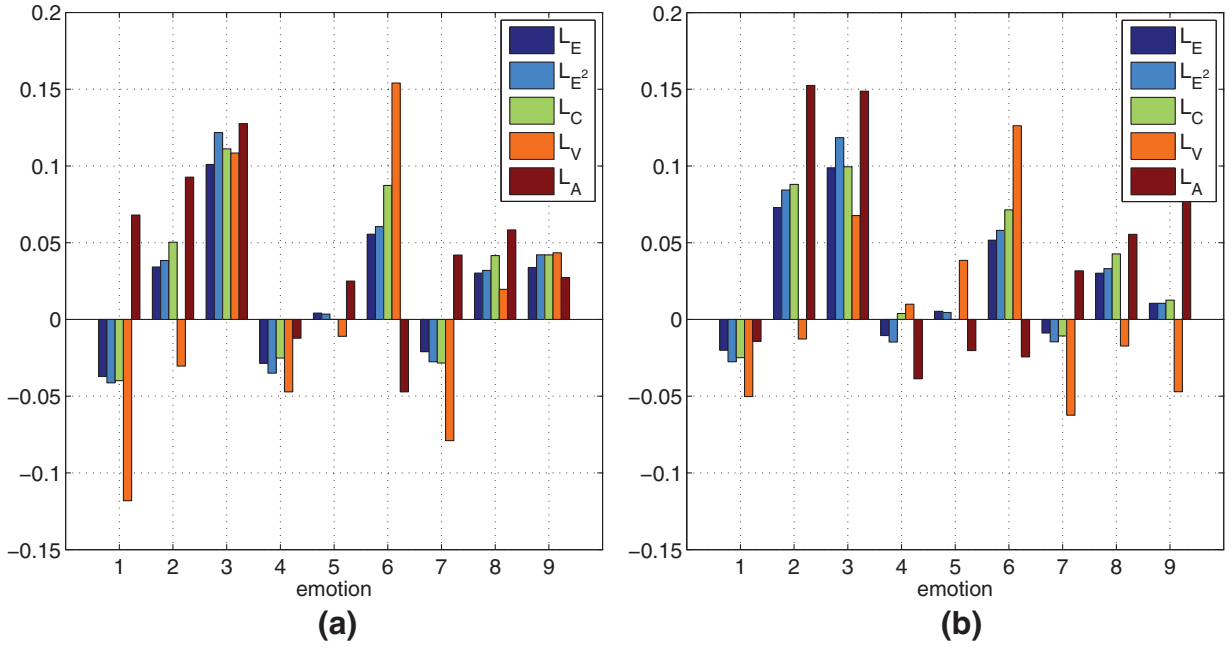

Fig. 11. Bar chart comparison of classification performances for (a) all participants combined, (b) average of individual participants. Neutral valence column: $e_{2,5,8}$; neutral arousal row: $e_{4,5,6}$.

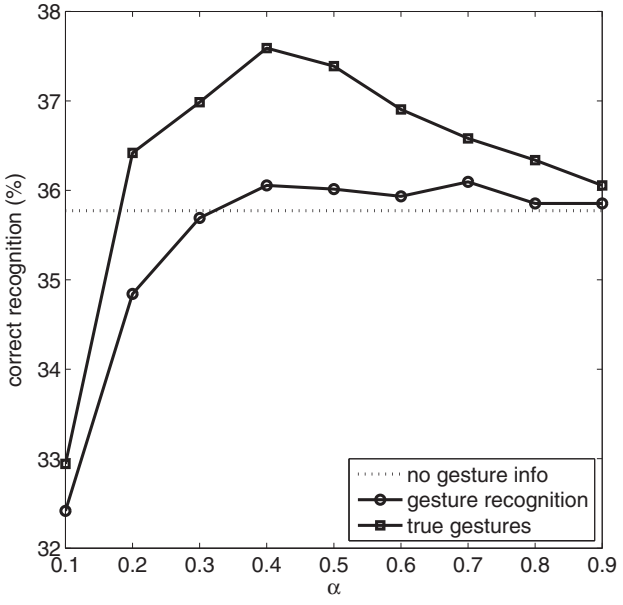

Fig. 12. Results combining gesture recognition with affect recognition.

participants to rate likelihood of performing a gesture for a given emotional state ([26], omitted for space). Prior probabilities were estimated by scaling likelihood values to produce a legitimate probability distribution.

We combined gesture information with affect recognition in two ways. First we combined them according to the gesture recognition performance of the random forest classifier, using Equation (1). Secondly, we combined the two results assuming that the gesture recognition was perfect. That is, we did not run the gesture recognition algorithm but considered the true labels for gestures and combined the results, manually setting $P_G(g_j|x) = 1$ for the true gesture and zero otherwise. Fig. 12 illustrates the results of the data fusion for different values of α. In this case, direct affect recognition performance without any gesture recognition aid is 35.9%. Combining random forest recognition of gestures results in a small performance increase. If we assume perfect gesture recognition, classification increases to 37.6% for $\alpha = 0.4$. This suggests that gesture recognition can indeed improve affect recognition results provided that gesture recognition performance is high. When our actual gesture recognition results are used, the improvement is marginal; 36.1% for $\alpha = 0.7$.

4. Conclusions and future work

In this study, we presented the results of affect recognition from touch gestures using FSR and accelerometer signals. The experiments were conducted with a lap-sized pet robot, the Haptic Creature. We considered Russell's 2-D arousal-valence affect space in modeling the emotions. The overall correct recognition rate is about 36% with all participants' data and 48% for individual participants on average. These numbers are comparable to the results from human-human affective touch studies (24%–63% and 31%–83% in two different studies) [12]. However, these human-human studies do not consider the closeness between emotions. While the obtained error rates are relatively high, we observed that in many cases, the misclassified states were close to the actual states in the affect space. To quantify and analyze this closeness, we defined several new metrics which demonstrate that it is possible to quantify and compare the misclassifications.

Knowledge of the nature of wrong classifications and near misses will be valuable in a deployed affective human-robot interaction system. For example, prior probabilities in a pattern recognition system can be adjusted according to near miss statistics, improving overall probability of correct classification. Depending on the application, asymmetric penalties may be associated with specific misclassifications. Knowing the performance of classifiers in arousal and valence dimensions would help improve the design of other classifiers. Integrating touch sensing with other sensing modalities (such as video or audio) would certainly improve results, another upcoming step.

We emphasize that our physical FSR coverage was low; there are regions on the surface where no sensor is present. We expect that better other technology will improve results considerably, and are updating sensing on a new platform.

Table 3
Confusion matrix for gestures. The highlighted diagonals indicates correct identification; mis-classification frequencies larger than the diagonal value are highlighted in red.

true class \ recognized class	CONTACT	CRADLE	FINGER IDLY	GRAB	HOLD	HUG	KISS	LIFT	MASSAGE	NUZZLE	PAT	PICK	POKE	PRESS	PULL	PUSH	ROCK	RUB	SCRATCH	SQUEEZE	STROKE	SWING	TAP	TICKLE	TOSS	TREMBLE
CONTACT	112	1	5	0	1	0	0	0	4	1	10	1	1	1	0	2	0	5	3	3	10	0	3	1	0	0
CRADLE	2	49	0	0	42	22	3	0	2	1	0	0	0	0	0	0	11	0	0	3	1	1	0	0	0	0
FINGER IDLY	8	1	30	0	0	0	0	0	6	0	12	2	0	0	0	0	0	12	5	1	30	0	3	12	0	0
GRAB	0	0	0	6	5	4	3	5	0	3	1	0	0	0	1	4	1	1	0	3	1	0	0	0	4	0
HOLD	0	37	1	1	42	34	8	3	1	6	2	0	0	0	8	0	6	1	0	1	2	0	0	0	0	0
HUG	2	18	0	0	10	113	2	2	0	2	0	0	0	1	2	0	0	1	0	2	1	0	0	0	0	0
KISS	4	2	1	0	9	4	25	4	0	8	1	0	0	1	0	0	5	0	0	2	0	0	0	0	1	0
LIFT	0	5	0	0	10	8	4	19	0	6	0	0	0	0	2	0	6	0	0	1	0	3	0	0	4	0
MASSAGE	3	2	6	0	2	1	0	0	23	0	2	1	0	0	0	0	2	55	10	1	12	0	1	0	0	0
NUZZLE	6	10	0	1	19	14	9	8	1	22	0	0	0	1	1	0	2	0	0	2	0	0	0	0	0	0
PAT	11	0	11	0	0	0	0	0	2	0	78	1	1	0	0	0	0	8	6	1	17	0	3	4	2	0
PICK	2	0	8	0	0	0	0	0	3	0	8	0	0	0	1	1	0	12	8	1	7	0	2	5	0	0
POKE	1	0	4	0	0	1	0	0	2	0	15	1	0	0	0	0	0	5	4	0	1	0	3	3	0	0
PRESS	17	0	1	0	0	0	0	0	4	0	3	0	0	4	0	5	0	9	0	6	1	0	1	0	0	0
PULL	3	2	0	2	11	4	1	2	1	3	0	0	0	0	19	3	0	1	0	9	0	0	0	0	1	0
PUSH	2	0	0	1	5	1	0	0	0	0	0	0	0	6	4	15	1	4	0	5	0	1	0	0	0	0
ROCK	1	30	0	0	13	6	0	1	2	1	0	0	0	0	0	1	28	5	1	1	0	4	0	0	1	1
RUB	6	0	6	0	1	0	0	0	29	0	3	1	0	1	0	0	1	56	9	2	38	0	0	3	0	1
SCRATCH	14	0	12	0	0	0	0	0	13	0	9	1	0	0	0	0	1	27	16	1	22	0	0	11	0	0
SQUEEZE	10	5	1	0	3	5	4	4	5	4	4	0	0	5	5	2	1	8	1	17	3	0	0	0	0	0
STROKE	14	2	17	0	0	0	0	0	9	0	16	0	0	0	0	0	0	37	9	1	78	0	0	1	0	0
SWING	0	7	0	0	9	7	0	3	0	4	0	0	0	0	0	0	5	0	0	0	0	10	0	1	7	0
TAP	11	0	4	0	0	0	0	0	0	0	31	2	0	0	0	0	0	2	1	0	4	0	5	4	0	1
TICKLE	6	2	19	0	2	0	0	0	1	0	9	5	0	0	1	0	0	12	14	1	18	0	1	9	0	0
TOSS	0	0	0	2	1	0	0	3	0	3	0	0	0	0	0	0	2	0	0	0	0	1	0	0	32	0
TREMBLE	2	1	1	0	3	2	0	0	1	1	2	0	0	0	0	0	5	7	5	0	1	0	1	0	0	2

In the following, we highlight other important results.

Location-independent features: We used location-independent features. This is appropriate for pure gesture recognition; however, location can provide useful information in *emotion* recognition. Here, adding location features did not improve results so we disregarded location. This finding could be due to the limited touch sensing capability of our robot; further investigation is necessary to identify the relation between touch location and emotion in human-pet interactions. On the other hand, our feature set does not totally disregard the location information. The accelerometer signal carries some information about the touch location, since, for example, a pat on the head and a pat on the back do not generate the same acceleration pattern.

Dimensional vs. categorical approach to affect: Our procedure began with a categorical approach to affect (we considered classes as independent); then, we interpreted results with a dimensional approach (affect grid), and show that misclassifications are usually "near misses," especially in the positive valence zone. This suggests that the 2-D arousal-valence grid is a good representation of affect in a human-robot interaction setting where touch is one of the sensing modalities.

Personalized affect recognition: With all participants' data combined, the best classification performance is a tie between e_3 (*excited*) and e_9 (*relaxed*). If near misses are taken into account, e_3 is correctly classified with higher performance than e_9. For the case with a single participant's data, e_2 (*aroused*) and e_5 (*neutral*) are correctly classified with the largest success. When near misses are taken into account, the classification performance of e_2 is better than that of e_5. This suggests that the affect classification performance for different emotions will be different for a pet robot that is used by many people as opposed to a personalized robot.

It is difficult to classify negative valence: Negative valence is harder to classify, especially in the medium-high arousal zone, when all emotional states are considered and with our sparse data conditions (three sensors out of 56). We also ran the algorithm with subsets of the emotional states; best results are obtained for emotions in the positive valence zone. It is possible that people are less likely to perform affective touch gestures on a pet robot in the negative valence emotional states, as compared to positive valence emotional states, but more data are needed to assess this. Our observation is confirmed by the numbers in the user survey previously reported in [26]; and Silvera-Tawil et al. [20] also report confusions mostly in negative valence and high arousal zone.

Affect recognition vs. gesture recognition: We combined gesture recognition information with direct affect recognition. For gesture recognition results obtained with the Random Forests method, there was little improvement when gesture data were added to the direct-affect classification. More improvement was achieved by assuming that the gesture recognizer was perfect. This suggests that good gesture recognition can be used to aid affect recognition systems based on touch sensing. Here we should state the fact that our method of estimating prior probabilities is based on empirical user survey data. Better ways of estimating prior probabilities and better gesture recognition methods could prove gesture recognition even more useful for affect recognition.

Human display of affect: In this study, participants were asked to enact an artificial display of affect. Recognizing naturalistic and spontaneous affect is still one of the problems of affect recognition systems that use audiovisual cues, but that problem applies to affective touch as well. We believe that our study is a stepping stone in solving this harder problem.

In this study, the touch recipient is a pet robot. Were the recipient a human or a humanoid robot, the human performing the touch would possibly display a particular emotion in a different manner—that is, the identity of the recipient may influence the touch

given. For this reason, our study should not be directly compared to studies in human-human(oid) interaction. Further investigation of human behavior in different human-human, human-robot, human-pet interactions will improve applications involving emotion recognition such as ours.

Future work

Affect recognition from touch gestures is currently in infant stages. To our knowledge, this study is one of the first in this area. There are many unexplored areas for future work. With the advance of sensor technology, cheap sensors that cover the entire touch surface can be built and spatially continuous data can be acquired. Finding the most informative features for affect recognition from touch gestures is another area where little research has been conducted. Unsupervised feature learning and deep learning approaches can be applied and better recognition performance could be achieved. Another question that needs to be answered is whether there is a need for recognizing gestures before recognizing emotional states. Further studies need to be conducted to answer these questions. We believe that video recordings of touch gestures provide additional information about the gesture being performed, as well as the emotional state. Considering the vast literature on affect recognition from vision, fusion of touch sensor and video data are another very promising method for affect recognition.

Acknowledgments

We thank our experiment participants, and Dr. M. Sedlmair for visualization advice. This work was supported by NSERC.

References

[1] J. Bottorff, The meaning of touch in caring for patients with cancer. Oncol. Nurs. Forum. 20 (1993) 1531–1538.
[2] L. Breiman, Random forests. Mach. Learn. 45 (2001) 5–32.
[3] R. A. Calvo, S. D'Mello, Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Trans. Affect. Comput. 1 (2010) 18–37.
[4] R. T. Clemen, R. L. Winkler, Combining probability distributions from experts in risk analysis. Risk Anal. 19 (1999) 187–203.
[5] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. G. Taylor, Emotion recognition in human-computer interaction. IEEE Signal Process. Mag. 18 (2001) 32–80.
[6] A. Flagg, K. E. MacLean, Affective touch recognition for a furry therapeutic machine, in: Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction, TEI'13, Barcelona, Spain, 2013, pp. 25–32.
[7] A. Flagg, D. Tam, K. E. MacLean, R. Flagg, Conductive fur sensing for a gesture-aware furry robot, in: Proceedings of IEEE Haptics Symposium, 2012, pp. 23–30.
[8] M. Fujita, AIBO: Toward the era of digital creatures. Int. J. Robot. Res. 20 (2001) 781–794.
[9] H. Güneş, M. Pantic, Automatic, dimensional and continuous emotion recognition. Int. J. Synth. Emotion 1 (2010) 68–99.
[10] H. Güneş, B. Schuller, M. Pantic, R. Cowie, Emotion representation, analysis and synthesis in continuous space: A survey, in: Proceedings of the 1st International Workshop on Emotion Synthesis, Representation, and Analysis in Continuous Space, IEEE FG 2011, Santa Barbara, CA, USA, 2011, pp. 827–834.
[11] M. J. Hertenstein, R. Holmes, D. Keltner, M. McCullough, The communication of emotion via touch. Emotion 9 (2009) 566–573.
[12] M. J. Hertenstein, D. Keltner, B. App, B. A. Bulleit, A. R. Jaskolka, Touch communicates distinct emotions. Emotion 6 (2006) 528–533.
[13] S. Jones, A. Yarbrough, A naturalistic study of the meanings of touch. Commun. Monogr. 52 (1985) 19–58.
[14] A. Montagu, Touching: The Human Significance of the Skin, third ed., Harper & Row, New York, NY, 1986.
[15] J. A. Russell, A circumplex model of affect. J. Pers. Soc. Psychol. 39 (1980) 1161–1178.
[16] J. A. Russell, A. Weiss, G. A. Mendelsohn, Affect grid: A single-item scale of pleasure and arousal. J. Pers. Soc. Psychol. 57 (1989) 493–502.
[17] J. Saldien, K. Goris, S. Yılmazyı ldız, W. Verhelst, D. Lefeber, On the design of the huggable robot Probo. J. Phys. Agent 2 (2008) 3–12.
[18] O. Schneider, K. E. MacLean, K. Altun, I. Karuei, M. Wu, Real-time gait classification for persuasive smartphone apps: Structuring the literature and pushing the limits, in: Proceedings of the International Conference on Intelligent User Interfaces, IUI 2013, Santa Monica, CA, USA, 2013, pp. 161–172.
[19] T. Shibata, K. Inoue, R. Irie, Emotional robot for intelligent system—Artificial emotional creature project, in: Proceedings of 5th IEEE International Workshop on Robot and Human Communication, Tsukuba, Ibaraki, Japan, 1996, pp. 466–471.
[20] D. Silvera-Tawil, D. Rye, M. Velonaki, Interpretation of social touch on an artificial arm covered with an EIT-based sensitive skin. Int. J. Soc. Robot. 6 (2014) 489–505.
[21] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, M. Wolf, The design of the Huggable: A therapeutic robotic companion for relational, affective touch, in: Proceedings of the AAAI Fall Symposium on Caring Machines: AI in Eldercare, The AAAI Press, Menlo Park, California, 2005, pp. 91–98.
[22] M. Wöllmer, F. Eyben, S. Reiter, B. Schuller, C. Cox, E. Douglas-Cowie, R. Cowie, Abandoning emotion classes—Towards continuous emotion recognition with modelling of long-range dependencies, in: Proceedings of Interspeech, Brisbane, Australia, 2008, pp. 597–600.
[23] S. Yohanan, The haptic creature: Social human-robot interaction through affective touch (Ph.D. thesis), The University of British Columbia, 2012.
[24] S. Yohanan, K. E. MacLean, A tool to study affective touch: Goals and design of the Haptic Creature, in: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, CHI'09, Boston, MA, USA, 2009, pp. 4153–4158.
[25] S. Yohanan, K. E. MacLean, Design and assessment of the Haptic Creature's affect display, in: Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction, HRI'11, Lausanne, Switzerland, 2011, pp. 473–480.
[26] S. Yohanan, K. E. MacLean, The role of affective touch in human-robot interaction: Human intent and expectations in touching the Haptic Creature. Int. J. Soc. Robot. 4 (2012) 163–180.
[27] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 39–58.