J Multimodal User Interfaces
DOI 10.1007/s12193-016-0232-9

ORIGINAL PAPER

Automatic recognition of touch gestures in the corpus of social touch

Merel M. Jung[1] · **Mannes Poel**[1] · **Ronald Poppe**[2] · **Dirk K. J. Heylen**[1]

Received: 12 August 2015 / Accepted: 13 October 2016

Abstract For an artifact such as a robot or a virtual agent to respond appropriately to human social touch behavior, it should be able to automatically detect and recognize touch. This paper describes the data collection of CoST: Corpus of Social Touch, a data set containing 7805 captures of 14 different social touch gestures. All touch gestures were performed in three variants: gentle, normal and rough on a pressure sensor grid wrapped around a mannequin arm. Recognition of these 14 gesture classes using various classifiers yielded accuracies up to 60 %; moreover, gentle gestures proved to be harder to classify than normal and rough gestures. We further investigated how different classifiers, interpersonal differences, gesture confusions and gesture variants affected the recognition accuracy. Finally, we present directions for further research to ensure proper transfer of the touch modality from interpersonal interaction to areas such as human–robot interaction (HRI).

Keywords Social touch · Touch corpus · Touch gesture recognition

✉ Merel M. Jung
m.m.jung@utwente.nl

Mannes Poel
m.poel@utwente.nl

Ronald Poppe
r.w.poppe@uu.nl

Dirk K. J. Heylen
d.k.j.heylen@utwente.nl

[1] University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

[2] University of Utrecht, P.O. Box 80125, 3508 TC Utrecht, The Netherlands

1 Introduction

Touch gestures can be used in social interaction to communicate and express different emotions [14]. For example, love can be communicated by hugging and stroking while anger can be expressed by pushing and shaking [15]. The sense of touch can also be used to explore our environment and manipulate objects such as tools, which can be highly functional [13]. As opposed to functional touch, Haans and IJsselsteijn [13] described social touch as all instances of interpersonal touch, whether this is accidental (e.g. bumping into someone on the street) or conscious (e.g. comforting someone who is upset). Here, we broaden this definition to include social touch interaction between humans and artifacts such as robots and virtual agents.

Extending social touch interaction to include interaction between humans and artifacts can result in more natural interaction, providing opportunities for various applications such as training medical students to use appropriate social touch behavior in a health care scenario involving a virtual patient [28]. Also, the addition of tactile interaction can benefit robot therapy in which robots are used to comfort people in stressful environments, for instance, children in hospitals [19] and elderly people in nursing homes [37].

We speak of social touch intelligence when an artifact, for example a robot, is able to understand the social meaning of human touch and is able to use touch in a socially appropriate way (see Fig. 1). In humans, receptors in the skin, muscles, joints and tendons register touch [11,34]. Equipping an artifact with touch sensors is the first step towards touch interaction based on human touch input. Once the sensor registers the touch, we need to recognize the type of touch and interpret its meaning. A robust touch recognition system should be perceived as working in real time. Also, gesture recognition should be subject independent to avoid training

Published online: 21 October 2016

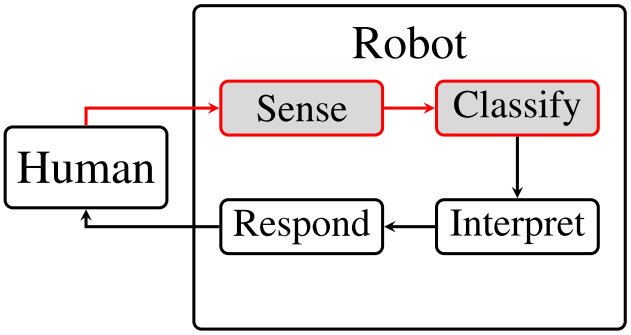

Fig. 1 Interaction cycle for a robot with social intelligence to respond to human touch. The steps that are the focus of this work are *highlighted*

sessions for new users. Some promising attempts have been made to recognize different sets of touch gestures (e.g. stroke, poke, and hit) recorded on various interfaces. As recognition rates vary depending on the degree of similarity between the touch gestures it is difficult to judge the relative strengths of one approach over the other.

To work towards reliable touch gesture recognition we have recorded a corpus of social touch hand gestures to characterize various touch gestures. We focus on the recognition of a list of relevant social touch gestures. The interpretation of the social meaning of these touch gestures is beyond the scope of this work. To the best of our knowledge there are no publicly available datasets on social touch for research and benchmarking. The contribution of this paper is three-fold: first, we will give a systematic overview of the characteristics of available studies on the recognition of social touch; second, we will give additional information about the Corpus of Social Touch (CoST); third, we will compare the performance of different classifiers to provide a baseline for touch gesture recognition within CoST and evaluate the factors that influence the recognition accuracy.

The remainder of the paper is organized as follows: the next section will discuss related work on the recognition of social touch, Sect. 3 will describe CoST. Touch gesture recognition results will be presented and discussed in Sects. 4 and 5, respectively. The paper will conclude in Sect. 6.

2 Related work on social touch recognition

There have been a number of studies on social touch recognition. These studies differ in their characteristics, which we will briefly discuss. A summary of previous studies is presented in Table 1. Please note that we have only considered the studies that reported details on classification and studies published up to August 2015.

2.1 Touch surface and sensors

In these studies, touch was performed on various surfaces such as robots (e.g. [26]), sensor sheets (e.g. [30]) or human body parts such as arms [32]. Physical appearances of interfaces for touch interaction included robotic animals (e.g. [39]), full body humanoid robots (e.g. [26]), partial embodiments such as a mannequin arm (e.g. [33]) and a balloon interface [29]. Several techniques were used for the sensing of touch, each having its own advantages and drawbacks for example, low cost vs. large hysteresis in force sensing resistors [8]. These sensing techniques were implemented in the form of artificial robot skins (e.g. [32]) or by following a modular approach using sensor tiles (e.g. [20]) or individual sensors to cover the robot's body (e.g. [5]). Designing an artificial skin entails extra requirements such as flexibility and stretchability to cover curved surfaces and moving joints [31,34] but has the advantage of providing equal sensor density for detection across the entire surface which can be hard to achieve using individual sensors [5]. The approach of using computer vision to register touch is noteworthy [7].

2.2 Touch recognition

Previous research on the recognition of touch has included hand gestures (e.g. stroke [23]), full body gestures (e.g. hug [6]), emotions (e.g. happiness [33]), and social messages (e.g. affection [33]). Data was gathered from a single subject to test a proof of concept (e.g. [5]) or from multiple subjects to allow for the training of a subject independent model (e.g. [33]). Classification results seem to show that it is harder to recognize emotions or social messages than the touch itself. This can be explained by the nontrivial nature of mapping touch to an emotional state or an intention for example, a single touch gesture can be used to communicate various emotions [15,39]. Also, as expected, results of a within-subjects design were better than classification between-subjects (e.g. [1]) meaning that there was a larger inter-person variance than intra-person variance. Human classification of touch out-performed automatic classification (e.g. [31]). However, when touch was mediated by technology, human performance decreased, Bailenson et al. [2] found that emotions were better recognized by participants when performing a real hand shake with another person compared to when the handshake with the other person was mediated through a force-feedback joystick. Classification was mostly offline however, some promising attempts have been made with real-time classification, which is a prerequisite for real-time touch interaction (e.g. [26]). Real-time systems come with extra requirements such as gesture segmentation and ensuring adequate processing speed. Combining computer vision with touch sensing yielded better touch recognition results than relying on a single modality [7].

Table 1 Results of literature on social touch recognition

Paper	Touch surface	Sensor(s)	Touch recognition of...	n	Classifier	Design	Accuracy
Altun and MacLean [1]	Haptic Creature	Force sensing resistors, accelerometer	26 Gestures	31	Random forest	Between-subjects	33 %
Altun and MacLean [1]	Haptic Creature	Force sensing resistors, accelerometer	9 Emotions	31	Random forest	Between-subjects	36 %
Altun and MacLean [1]	Haptic Creature	Force sensing resistors, accelerometer	9 Emotions	31	Random forest	Within-subjects	48 %
Altun and MacLean [1]	Haptic Creature	Force sensing resistors, accelerometer	9 Emotions using gesture recog.	31	Random forest	Between-subjects	36 %
Bailenson et al. [2]	Force-feedback joystick	2d accelerometer	7 Emotions	16	Classification by human	1 Subject rates 1 other	33 %
Bailenson et al. [2]	Force-feedback joystick	2d accelerometer	7 Emotions	16	SVM[a] RBF[b] kernel	Between-subjects	36 %
Bailenson et al. [2]	Other subject's hand	/	7 Emotions	16	Classification by human	1 Subject rates 1 other	51 %
Chang et al. [5]	Haptic Creature	Force sensing resistors	4 Gestures	1	Custom recognition software	Real-time	Up to 77 %
Cooney et al. [6]	Sponge (humanoid) robot	Accelerometer, gyro sensor	13 Full-body gestures	21	SVM[a] RBF[b] kernel	Between-subjects	77 %
Cooney et al. [7]	Humanoid robot 'mock-up'	Photo-interrupters	20 Full-body gestures	17	k-NN[c]	Between-subjects	63 %
Cooney et al. [7]	Humanoid robot 'mock-up'	Photo-interrupters	20 Full-body gestures	17	SVM[a] RBF[b] kernel	Between-subjects	72 %
Cooney et al. [7]	Humanoid robot 'mock-up'	Microsoft Kinect	20 Full-body gestures	17	k-NN[c]	Between-subjects	67 %
Cooney et al. [7]	Humanoid robot 'mock-up'	Microsoft Kinect	20 Full-body gestures	17	SVM[a] RBF[b] kernel	Between-subjects	78 %
Cooney et al. [7]	Humanoid robot 'mock-up'	Photo-interrupters, Microsoft Kinect	20 Full-body gestures	17	k-NN[c]	Between-subjects	82 %
Cooney et al. [7]	Humanoid robot 'mock-up'	Photo-interrupters, Microsoft Kinect	20 Full-body gestures	17	SVM[a] RBF[b] kernel	Between-subjects	91 %
Flagg et al. [9]	Furry lap pet	Conductive fur sensor, piezoresistive fabric pressure sensor	9 Gestures	16	Neural network	Between-subjects	75 %
Flagg et al. [9]	Furry lap pet	Conductive fur sensor, piezoresistive fabric pressure sensor	9 Gestures	16	Logistic regression	Between-subjects	72 %
Flagg et al. [9]	Furry lap pet	Conductive fur sensor, piezoresistive fabric pressure sensor	9 Gestures	16	Bayes network	Between-subjects	68 %
Flagg et al. [9]	Furry lap pet	Conductive fur sensor, piezoresistive fabric pressure sensor	9 Gestures	16	Random forest	Between-subjects	86 %
Flagg et al. [9]	Furry lap pet	Conductive fur sensor, piezoresistive fabric pressure sensor	9 Gestures	16	Random forest	Within-subjects	94 %
Flagg et al. [10]	Fur sensor	Conductive fur sensor	3 Gestures	7	Linear regression	Between-subjects	82 %

Table 1 continued

Paper	Touch surface	Sensor(s)	Touch recognition of...	n	Classifier	Design	Accuracy
Ji et al. [20]	KASPAR (hand section)	Capacitive pressure sensors	4 Gestures	1	SVM[a] intersection kernel	Within-subject	Up to 96 %
Ji et al. [20]	KASPAR (hand section)	Capacitive pressure sensors	4 Gestures	1	SVM[a] RBF[b] kernel	Within-subject	Up to 93 %
Jung [22]	Mannequin arm	Piezoresistive fabric pressure sensors	14 Gestures	31	Bayesian classifier	Subject-independent	53 %
Jung [22]	Mannequin arm	Piezoresistive fabric pressure sensors	14 Gestures	31	SVM[a] linear kernel	Subject-independent	46 %
Jung et al. [23]	Mannequin arm	Piezoresistive fabric pressure sensors	14 Rough gestures	31	Bayesian classifier	Subject-independent	54 %
Jung et al. [23]	Mannequin arm	Piezoresistive fabric pressure sensors	14 Rough gestures	31	SVM[a] linear kernel	Subject-independent	53 %
Kim et al. [26]	KaMERo	Charge-transfer touch sensors, accelerometer	4 Gestures	12	Temporal decision tree	Real-time	83 %
Knight et al. [27]	Sensate bear	Electric field sensor, capacitive sensors	4 Gestures	11	Bayesian networks + k-NN[c]	Real-time	20–100 %
Nakajima et al. [29]	Emoballoon	Barometric pressure sensor, microphone	6 Gestures + 'no touch'	9	SVM[a] RBF[b] kernel	Between-subjects	75 %
Nakajima et al. [29]	Emoballoon	Barometric pressure sensor, microphone	6 Gestures + 'no touch'	9	SVM[a] RBF[b] kernel	Within-subjects	84 %
Naya et al. [30]	Sensor sheet	Pressure-sensitive conductive ink	5 Gestures	11	k-NN[c] + Fisher's linear discriminant	Between-subjects	87 %
Silvera-Tawil et al. [31]	Sensor sheet	Pressure sensing based on EIT[d]	6 Gestures	1	Logitboost algorithm	Within-subject	91 %
Silvera-Tawil et al. [31]	Sensor sheet	Pressure sensing based on EIT[d]	6 Gestures	35	Logitboost algorithm	Between-subjects	74 %
Silvera-Tawil et al. [31]	Experimenter's back	/	6 Gestures	35	Classification by human	Between-subjects	86 %
Silvera-Tawil et al. [32]	Mannequin arm	Pressure sensing based on EIT[d], force sensor	8 Gestures + 'no touch'	2	Logitboost algorithm	Within-subjects	88 %
Silvera-Tawil et al. [32]	Experimenter's arm	/	8 Gestures	2	Classification by human	Within-subjects	75 %
Silvera-Tawil et al. [32]	Mannequin arm	Pressure sensing based on EIT[d], force sensor	8 Gestures + 'no touch'	40	Logitboost algorithm	Subject-independent	71 %
Silvera-Tawil et al. [32]	Other subject's arm	/	8 Gestures	40	Classification by human	1 Subject rates 1 other	90 %
Silvera-Tawil et al. [33]	Mannequin arm	Pressure sensing based on EIT[d], force sensor	6 Emotions + 'no touch'	2	Logitboost algorithm	Within-subjects	88 %
Silvera-Tawil et al. [33]	Mannequin arm	Pressure sensing based on EIT[d], force sensor	6 Social messages + 'no touch'	2	Logitboost algorithm	Within-subjects	84 %
Silvera-Tawil et al. [33]	Mannequin arm	Pressure sensing based on EIT[d], force sensor	6 Emotions + 'no touch'	2	Logitboost algorithm	Between-subjects	32 %

Table 1 continued

Paper	Touch surface	Sensor(s)	Touch recognition of...	n	Classifier	Design	Accuracy
Silvera-Tawil et al. [33]	Mannequin arm	Pressure sensing based on EIT^d, force sensor	6 Social messages + 'no touch'	2	Logitboost algorithm	Between-subjects	51 %
Silvera-Tawil et al. [33]	Mannequin arm	Pressure sensing based on EIT^d, force sensor	6 Emotions + 'no touch'	42	Logitboost algorithm	Subject-independent	47 %
Silvera-Tawil et al. [33]	Other subject's arm	/	6 Emotions	42	Classification by human	1 Subject rates 1 other	52 %
Silvera-Tawil et al. [33]	Mannequin arm	Pressure sensing based on EIT^d, force sensor	6 Social messages + 'no touch'	42	Logitboost algorithm	Subject-independent	50 %
Silvera-Tawil et al. [33]	Other subject's arm	/	6 Social messages	42	Classification by human	1 Subject rates 1 other	62 %
Stiehl et al. [35]	The Huggable (arm section)	Electric field sensor, force sensors, thermistors	8 Gestures (disregarding 'slap')	1	Neural network	Within-subject	79 %
van Wingerden et al. [38]	Mannequin arm	Piezoresistive fabric pressure sensors	14 Rough gestures	31	Neural network	Between-subjects	64 %

SVM support vector machine, *RBF* radial basis function, *k-NN* k-nearest neighbor, *EIT* electrical impedance tomography

Fig. 2 Participant performing the instructed touch gesture on the pressure sensor (*the black fabric*) wrapped around the mannequin arm

Direct comparison of touch recognition between studies based on reported accuracies is difficult because of differences in the number and nature of touch classes, sensors, and classification protocols. Furthermore, some reported accuracies were the result of a best-case scenario intending to be a proof of concept (e.g. [5]). Some studies focused on the location of the touch rather than the touch gesture, such as distinguishing between 'head-pat' and 'foot-rub' [27]. While information on body location can enhance touch recognition, Silvera-Tawil et al. showed that comparable accuracies can be achieved by limiting the touch location to a single arm [32].

3 CoST: corpus of social touch

To address the need for social touch datasets, we recorded a corpus of social touch gestures (CoST) which was introduced in [23]. This data set is publicly available [25].

3.1 Touch gestures

CoST consists of the pressure sensor data of 14 different touch gestures performed on a sensor grid wrapped around a mannequin arm (see Fig. 2). The touch gestures (see Table 2) included in the data collection were chosen from a touch dictionary composed by [39] based on the literature on touch interaction between humans and between humans and animals. The list of gestures was adapted to suit interaction with a mannequin arm. Touch gestures involving physical movement of the arm itself such as lift, push and swing, were omitted because the movement of the mannequin arm could not be sensed by the pressure sensors. All touch gestures were performed in three variants: gentle, normal and rough

Table 2 Touch dictionary, adapted from Yohanan and MacLean [39]

Gesture label	Gesture definition
Grab	Grasp or seize the arm suddenly and roughly
Hit	Deliver a forcible blow to the arm with either a closed fist or the side or back of your hand
Massage	Rub or knead the arm with your hands
Pat	Gently and quickly touch the arm with the flat of your hand
Pinch	Tightly and sharply grip the arm between your fingers and thumb
Poke	Jab or prod the arm with your finger
Press	Exert a steady force on the arm with your flattened fingers or hand
Rub	Move your hand repeatedly back and forth on the arm with firm pressure
Scratch	Rub the arm with your fingernails
Slap	Quickly and sharply strike the arm with your open hand
Squeeze	Firmly press the arm between your fingers or both hands
Stroke	Move your hand with gentle pressure over arm, often repeatedly
Tap	Strike the arm with a quick light blow or blows using one or more fingers
Tickle	Touch the arm with light finger movements

to increase the variety of ways a gesture could be performed by each individual.

3.2 Pressure sensor grid

For the sensing of the gestures, an 8 × 8 pressure sensor grid (PW088-8x8/HIGHDYN from *plug-and-wear*[1]) was connected to a Teensy 3.0 USB Development Board (by *PJRC*[2]). The sensor was made of textile consisting of five layers. The two outer layers were protective layers made of felt. Each outer layer was attached to a layer containing eight strips of conductive fabric separated by non-conductive strips. Between the two conductive layers was the middle layer which comprised a sheet of piezoresistive material. The conductive layers were positioned orthogonally so that they formed an 8 by 8 matrix. The sensor area was 160 × 160 mm with a thickness of 4 mm and a spatial resolution of 20 mm.

One of the conductive layers was attached to the power supply while the other was attached to the A/D converter of the Teensy board. After A/D conversion, the sensor values of the 64 channels ranged from 0 to 1023 (i.e., 10 bits). Figure 3 displays the relationship between the sensor values and the pressure in kg/cm^2 for both the whole range (0–1023) and the range used in the data collection (0–990). Pressure used during human touch interaction typically ranges from 30 to 1000 g/cm^2 [34], which corresponds to sensor values between 25 and 800. From the plots it can be seen that the sensor's resolution is accurate within this range but decreases at higher pressure levels. Sensor data was sampled at 135 Hz.

Our sensor meets the requirements set by Silvera-Tawil et al. [34] for optimal touch sensing in social human–robot interaction as the spatial resolution falls within the recommend range of 10–40 mm and the sample rate exceeds the required minimum (20 Hz). However, the human somatosensory system is more complex than this sensor as receptors in the skin register not only pressure but also pain and temperature and receptors in the muscles, joints and tendons register body motion [11,34]. The sensor grid produces artifacts in the signal such as crosstalk, wear out and hysteresis (i.e., the influence of the previous and current input, which is discussed in Sect. 3.4). For demonstration purposes, we illustrated the sensor's crosstalk by pushing down with the back of a pencil perpendicular to the sensor grid to create a concentrated load (see Fig. 4). The sensor was wrapped around the mannequin arm to create a setup similar to the one used for the data collection. We did not compensate for the artifacts in the data.

3.3 Data acquisition

3.3.1 Setup

The sensor was attached to the forearm of a full size rigid mannequin arm consisting of the left hand and the arm up to the shoulder (see Fig. 2). The arm was chosen as the contact surface because this is one of the body parts where emotions can be communicated [15]. Also, the arm is one of the least invasive body areas on which to be touched [16] and presumably a neutral body location to touch others. The mannequin arm was fastened to the right side of the table to prevent it from slipping. Instructions for which gesture to perform had been scripted using *PsychoPy*[3] and were displayed to the participants on a computer monitor. Video recordings were made during the data collection as verification of the sensor data and the instructions given.

3.3.2 Procedure

Upon entering the experimenter room, the participant was welcomed and was asked to read and sign an informed consent form. After filling in demographic information, the participant was provided with a written explanation of the data collection procedure. Participants were instructed to use their right hand to perform the touch gestures and use their left

[1] www.plugandwear.com.

[2] www.pjrc.com.

[3] A module written for Python, see www.psychopy.org.

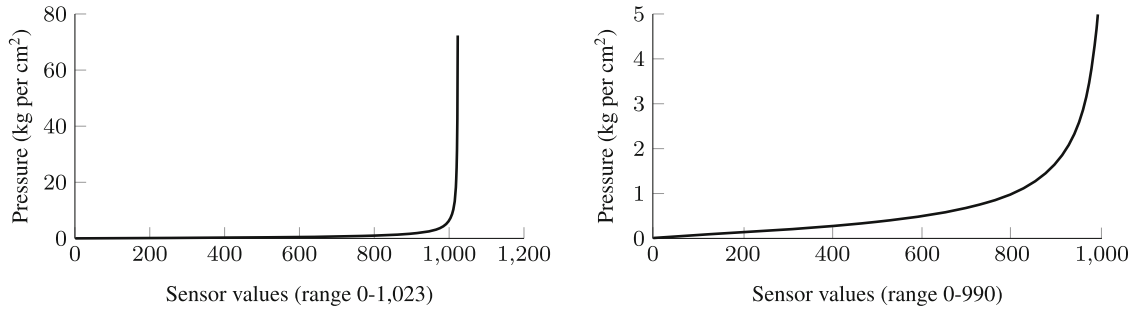

Fig. 3 Plot of the relationship between the sensor output after A/D conversion and pressure in kg/cm^2 for both the whole range (*left*) and the range used (*right*)

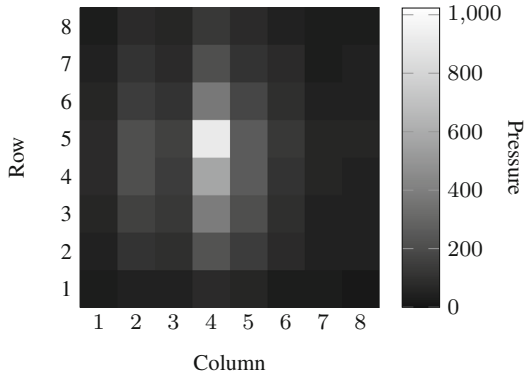

Fig. 4 Crosstalk visualization showing the sensor data of a single frame, a pencil was pressed down on the sensor grid (light pressure point) effecting the pressure level of adjacent channels

hand on the keyboard. Then an instruction video was shown of a person performing all 14 gestures on the mannequin arm based on the definitions from Table 2. Participants were instructed to repeat every gesture from the video to practice. No video examples were shown during the actual data collection. Thereafter example instructions were given to perform a stroke gesture in all three variants (i.e., gentle, normal and rough). After each gesture the participant could press the *spacebar* to continue to the next gesture or *backspace* to retry the current gesture. Once everything was clear to the participant the data collection started.

During the data collection each participant was prompted with 14 different touch gestures 6 times in 3 variants resulting in 252 gesture captures. In the instructions of the gesture to perform, the participants were shown only the gesture variant combined with the name of the gesture (e.g. 'gentle grab'), not the definition from Table 2. The order of instructions was pseudo-randomized into three blocks. Each instruction was given two times per block but the same instruction was not given twice in consecutive order. A single fixed list of instructions was constructed using these criteria. This list and the reversed order of the list were used as instructions in a counterbalanced design. After each block, there was a break and the participant was asked to report any difficulty in performing the instructions. Finally, participants were asked to give their own definitions of the gestures and manners. The entire procedure took approximately 40 minutes for each participant.

3.3.3 Participants

A total of 32 people volunteered to participate in the data collection. Data of one participant was omitted due to technical difficulties. The remaining participants, 24 male and 7 female, all studied or worked at the University of Twente in the Netherlands. Most (26) had the Dutch nationality (1 British/Dutch), others were Ecuadorean, Egyptian, German (2×) and Italian. The age of the participants ranged from 21 to 62 years ($M = 34$, $SD = 12$) and 29 were right-handed.

3.4 Data preprocessing

The raw data was segmented into gesture captures based on the key strokes of the participants marking the end of a gesture. Segmentation between keystrokes still contained many additional frames from before and after the gesture was performed. Removing these additional frames is especially important to reduce noise in the calculation of features that contain a time component, such as features that average over frames in time. See Fig. 5 for an example of a gesture capture of 'normal tap' as segmented between keystrokes, further segmentation is indicated by dashed lines. This plot also illustrates that the sensor values remain non-zero (the absolute minimum) when the sensor was not touched and that hysteresis occurs, in this case the sensor values are higher after the touch gesture is performed compared to before.

Further segmentation of the gesture captures was based on the change in the gesture's intensity (i.e., the summed pressure over all 64 channels) over time using a sliding window approach. The first window starts at the beginning of the gesture capture and includes the number of frames corresponding to the window size parameter. The next window

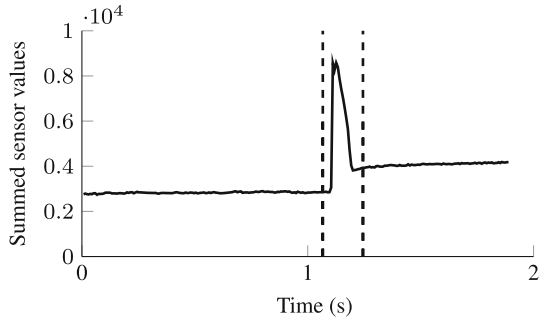

Fig. 5 Gesture capture of 'normal tap' as segmented between keystrokes, further segmentation based on pressure difference is indicated by the *dashed lines*

remains the same size but is shifted a number of frames corresponding to the step size parameter. The pressure intensity of each window is compared to that of the previous window. This procedure continues till the end of the gesture capture. Parameters (i.e., threshold of minimal pressure difference, step size, window size and offset) were optimized by visual inspection to ensure that all gestures were captured within the segmented part. The optimized parameters were fixed for all recordings.

After visual inspection it turned out that six gesture captures could not be automatically segmented because differences in pressure were too small (i.e., below the threshold parameter). The video recordings revealed that the gestures were either skipped or were performed too fast to be distinguishable from the sensor's noise. One other gesture capture was of notably longer duration (over a minute) than all other instances because the instructions were unclear at first. These seven gesture captures were instances of the variants 'gentle massage', 'gentle pat', 'gentle stroke', 'normal squeeze', 'normal tickle', 'rough rub', and 'rough stroke'. The instances of these gesture variants were removed from the dataset. The remaining dataset consists of 7805 touch gesture captures in total: 2601 gentle, 2602 normal and 2602 rough gesture captures.

3.5 Descriptive statistics

To get an idea of the differences between touch gestures and the variants, descriptive statistics were calculated on three important characteristics of touch: intensity (g/cm^2), contact area (% of sensor area) and gesture duration (ms). Pressure intensity was calculated as the mean pressure of all channels averaged over time and the maximum channel value of the gesture over all channels. Contact area was calculated for the frame with the highest summed pressure over all channels (corresponds to feature 21). Means and standard deviations of the touch data after segmentation are displayed for each variant and in total in Table 3 and per gesture in Table 4. It is notable that the mean and maximum pressure used per variant follow the expected pattern: gentle variants $<$ normal variant $<$ rough variants, indicating that participants used pressure to distinguish between the different variants. Figure 6 illustrates that there was a lot of overlap in duration between the different gestures (e.g. between hit and slap) and a lot of variance within each gesture, especially within massage and tickle. The tables and figure illustrate that the challenge of touch gesture recognition is complex and that it is not possible to distinguish between these different touch gestures using only these descriptive statistics. Table 5 shows the touch characteristics for males and females separately. Based on these characteristics there seems to be no significant differences between male and female touch gestures.

3.6 Self reports

In the self reports the most common difficulties (mentioned by at least 5 out of 31 participants) of distinguishing between gestures (disregarding the variants) were reported on pat vs. tap (12), grab vs. squeeze (10), rub vs. stroke (7), hit vs. slap (5) and pinch vs. squeeze (5). Furthermore, some combinations of gestures with variants were perceived as less logical. The most commonly mentioned gesture variants were: rough tickle (4), gentle hit (3) and gentle slap (3). Also, three participants raised concerns about breaking the setup when performing gestures too roughly.

At the end of the experiment participants were asked to provide their own definitions. The most common keywords used to define the gentle variant were: soft (mentioned by 8 participants), slow (6), less force (6), less pressure (5), and light (3) while the rough variant was defined as: more force (12), hard (7), more pressure (4), and fast (3), energetic (3). 'Normal' was defined as: the default/ regular (7), without thinking (4) and neutral (3).

Table 3 Mean and standard deviation (in parentheses) of the duration, mean and maximum pressure and contact area per touch variant and for all data

Variant	Gentle	Normal	Rough	All
Mean pressure (g/cm^2)	115 (61)	136 (82)	189 (157)	147 (112)
Max pressure (g/cm^2)	894 (511)	1260 (629)	1983 (813)	1379 (802)
Contact area (% of sensor)	.21 (.16)	.22 (.18)	.26 (.21)	.23 (.19)
Duration (ms)	1385 (1303)	1377 (1257)	1500 (1351)	1421 (1305)

Table 4 Mean and standard deviation (in parentheses) of the duration (ms), mean and maximum pressure (g/cm^2) the contact area (% of sensor area) per touch gesture

Gesture	Grab	Hit	Massage	Pat	Pinch	Poke	Press	Rub	Scratch	Slap	Squeeze	Stroke	Tap	Tickle
Mean pressure	349 (191)	101 (32)	172 (77)	100 (33)	126 (45)	95 (27)	188 (99)	131 (45)	106 (28)	95 (30)	286 (180)	116 (35)	92 (30)	96 (26)
Max pressure	1774 (919)	1643 (854)	1621 (800)	1057 (568)	1701 (892)	1258 (793)	1660 (802)	1282 (671)	1064 (524)	1165 (557)	1980 (946)	1135 (623)	1055 (610)	911 (497)
Contact area	.59 (.17)	.15 (.05)	.36 (.20)	.15 (.07)	.12 (.08)	.08 (.08)	.23 (.16)	.21 (.10)	.17 (.08)	.15 (.06)	.47 (.24)	.20 (.08)	.12 (.07)	.18 (.09)
Duration	1373 (715)	337 (403)	3538 (1898)	709 (753)	1132 (597)	650 (502)	1181 (608)	2170 (1142)	2205 (1268)	321 (462)	1502 (813)	1722 (829)	564 (486)	2491 (1446)

4 Recognition of social touch gestures

In this section we will present the performance results of several classifiers for the recognition of touch gestures in CoST. To establish the benchmark performance for CoST we compared the performance of four different commonly used classifiers. Two simple classifiers were chosen: a statistical model (Bayesian classifier) and a decision tree which allows for more insight into the classification process (e.g. which features are most important). Furthermore, we chose two more complex classifiers: a Support Vector Machine (SVM) which uses a single decision boundary and a neural network which allows for more complex decision boundaries.

4.1 Feature extraction

The data from the pressure sensor consists of a pressure value (i.e., the intensity) per channel (i.e., the location) at 135 fps. From the recorded sensor data, features were extracted for every gesture capture. The majority of features were based on the literature. The first features (1–28) were taken from previous work on this data set [22,23] which were based on social touch recognition literature, differences are indicated. Features used for video classification can be applied to this data because the data of CoST is a grid of pressure values that are updated at a fixed rate which is similar to a low-resolution gray scale video. Features 29–43 were slight adaptations of the features used in [38] which were based on video classification literature. Feature numbers are indicated in parentheses.

- *Mean pressure* is the mean over channels and time (1).
- *Maximum pressure* is the maximum value over channels and time (2).
- *Pressure variability* is the mean over time of the sum over channels of the absolute value of difference between two consecutive frames (3).
- *Mean pressure per row* is the mean over columns and time resulting in one feature per row which are in the direction of the mannequin arm's length (from top to bottom, 4–11).
- *Mean pressure per column* is the mean over rows and time resulting in one feature per column which are in the direction of the mannequin arm's width (from left to right, 12–19).
- *Contact area* per frame is the fraction of channels with a value above 50 % of the maximum value. Mean contact area is the mean over time of contact area (20) and the maximum pressure contact area is the contact area of the frame with the highest mean pressure over channels (21). The size of the contact area indicated whether the whole hand was used for a touch gesture, as would be expected

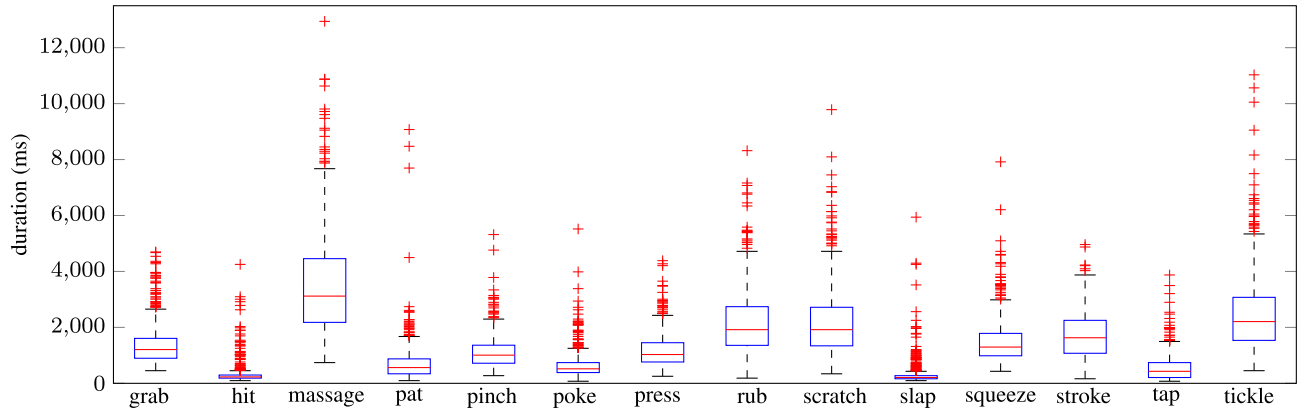

Fig. 6 *Boxplot* of the duration (ms) for all data per touch gesture

Table 5 Mean and standard deviation (in parentheses) of the duration, mean and maximum pressure and contact area per touch variant and for all data for male and female subjects

Variant	Gentle	Normal	Rough	All
Male				
Mean pressure (g/cm^2)	117 (63)	137 (85)	193 (163)	149 (117)
Max pressure (g/cm^2)	885 (518)	1245 (629)	1981 (828)	1370 (811)
Contact area (% of sensor)	.21 (.16)	.22 (.18)	.27 (.22)	.23 (.19)
Duration (ms)	1358 (1296)	1349 (1249)	1491 (1357)	1399 (1303)
Female				
Mean pressure (g/cm^2)	112 (50)	130 (72)	175 (133)	139 (96)
Max pressure (g/cm^2)	925 (485)	1310 (624)	1990 (763)	1409 (772)
Contact area (% of sensor)	.20 (.15)	.21 (.17)	.24 (.20)	.21 (.17)
Duration (ms)	1477 (1325)	1476 (1281)	1528 (1330)	1494 (1312)

for grab, or for example only one finger, as would be expected for a poke.

– *Temporal peak count* indicated how many times there was a significant increase in pressure level that is, whether a touch gesture consisted of continuous touch contact as would be expected for grab or alternating pressure levels which would be expected for a tickle. One feature counts the number of frames for which the average pressure of a frame was larger than that of its neighboring frames (22). (This feature replaced the previous version of features 22 from [22,23]). The other feature was calculated as the number of positive crossings of the threshold. The threshold was the mean over time of the pressure summed over all channels (23).
– *Traveled distance* (previously called 'displacement' in [22,23]) indicated the amount of movement of the hand across the contact area. For example, for a squeeze less movement across the sensor grid would be expected than for a stroke. Center of mass (i.e., the average channel weighted by pressure) was used to calculate the movement on the contact surface in both the row and column direction. Two features were calculated in the row direction: the mean traveled distance of the center of mass over time (24) and the summed absolute difference of the center of mass over time (25). The same features were calculated for the column direction (26–27).
– *Duration* of the gesture measured in frames (28).
– *Pressure distribution* (previously called 'histogram-based features' in [38]) is the normalized histogram over all channels and time of the pressure values. The histogram contains eight bins equally spaced between 0 and 1023 (29–36).
– *Spatial peaks* (previously called 'motion-based features' in [38]). Spatial peaks A spatial peak in a frame is a local maximum with a value higher than 0.75 of the the maximum pressure (see feature 2). The following features were derived from the local peaks; the mean (37) and variance (38) over time of the number of spatial peaks per frame. Also the mean over all spatial peaks and time of the distance of the spatial peak to the center of mass is a feature (39). The last feature based on spatial peaks is the mean over time and spatial peaks of the chance in distance of each peak w.r.t. the center of mass (40).
– *Derivatives* were calculated as the mean absolute pressure differences within the rows and columns between frames. Features were derived from the mean over time

and rows or columns of the above values (41–42). Also the mean absolute pressure difference for all channels was calculated. The last feature was based on the mean over time and channels (43).
- *Variance* over channels and time (44).
- *Direction of movement* indicated the angle in which the center of mass was moving between frames. These angle values were divided into quadrants of 90° each. For example, if the hand moves from the middle of the sensor grid to the upper right corner of the sensor grid, the center of mass moves at a 45° angle which falls within the upper right quadrant (i.e., the first quadrant). To deal with vectors that were close to the edge of two quadrants two points around the vector were evaluated, each weighting 0.5. A histogram represented the percentage of frames that fell into each quadrant (45–48).
- *Magnitude of movement* indicated the amount of movement of the center of mass. Statistics on the magnitude were calculated per gesture consisting of the mean, standard deviation, sum, and the range (49–52).
- *Periodicity* was the frequency with the highest amplitude in the frequency spectrum of the movement of the center of mass in the row and column direction, respectively (53–54).

4.2 Classification experiments

The extracted features were used for classification in MATLAB® (release 2013a). We performed two classification experiments: (1) classification of the touch gestures from the total dataset based on the gestures' class, thereby disregarding the variant (e.g. 'gentle grab' and 'normal grab' both belong to the same class: 'grab'); (2) classification of the touch gestures within each variant, splitting the data into three subsets: normal, gentle and rough. Due to their more pronounced nature, rough gesture variants were expected to have a more favorable signal-to-noise ratio compared to the softer variants.

For both classification experiments the data was split into a train/ validation set and test set using leave-one-subject-out cross-validation (31 folds) to train a user-independent model (i.e., data from each subject was only part of either the train set or the test set). Hyperparameters were optimized on the train/validation set using leave-one-subject-out cross-validation (30 folds). Classification results were evaluated using the best performing hyper parameters found from the 30 folds (i.e., training/validation set only) to classify the test set. This procedure was repeated for all 31 folds. Note that each fold can have different optimized parameter values. The baseline of classifying a sample into the correct class based on random guessing is $1/14 \approx 7$ % for both experiments. We will discuss details of each of the classifiers individually.

4.2.1 Bayesian classifier

The Gaussian Bayesian classifier has no hyperparameters to optimize. The mean and covariance for the features per class were calculated from the training data. These parameters for the multivariate normal distribution were used to calculate the posterior probability of a test sample belonging to the given class. Samples were assigned to the class with the maximum posterior probability.

4.2.2 Decision tree

Decision trees were trained using the CART learning algorithm with Gini's diversity index as splitting method. First a full tree was grown after which the tree was pruned. A parameter search for the optimal pruning level, using cross-validation as described above, was performed using a range of 5–30 in increments of 5.

4.2.3 Support vector machine

SVMs were trained using the LIBSVM software library [4], both with a linear kernel (hyper parameter C) and with a Radial Basis Function (RBF) kernel (hyper parameters C and γ). We chose to test two kernels due to their different approaches, the linear kernel separates the classes globally while the RBF kernel allows for a local division of two classes. The hyperparameters were optimized, using cross-validation as described above. A (grid)search was conducted for optimal parameters by growing the sequences of the parameter values exponentially ($C = 2^{-5}, 2^{-3}, \ldots, 2^{15}; \gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$) as proposed by [17]. Before training, features were rescaled to the range of [0, 1] by subtracting the minimum feature value from all feature values and dividing the result by the range of the feature values. Scaling prevents features with greater numeric ranges from dominating those with smaller numeric ranges [17].

4.2.4 Neural network

A feedforward neural network was trained using Levenberg–Marquardt optimization. Stopping criteria were set to a maximum of 1000 training iterations or six subsequent increases of the error on the validation set. The neural network toolbox in MATLAB automatically maps the range of the original input features to the range of $[-1, 1]$. Because of memory constraints the architecture was set to two layers of 54 and 27 neurons, respectively to get results in a timely fashion. Leave-one-subject-out cross-validation was used, the data from the remaining 30 subjects was split into a train set (70 %) and a validation set (30 %). The best performing network on the validation set of five runs was used to evaluate the test set (i.e., the samples of the left-out subject).

Table 6 Overall accuracies of leave-one-subject-out cross-validation for the variants per classifier, standard deviations in parentheses

	Variant			
	All	Normal	Gentle	Rough
Classifier				
Bayesian	.57 (.11)	.59 (.13)	.52 (.14)	.58 (.12)
Decision tree	.48 (.10)	.49 (.13)	.43 (.10)	.52 (.10)
SVM linear	.59 (.11)	.60 (.11)	.54 (.13)	.62 (.13)
SVM RBF	.60 (.11)	.60 (.11)	.54 (.13)	.62 (.12)
Neural network	.59 (.12)	.58 (.13)	.52 (.13)	.59 (.13)

4.3 Results

Table 6 provides an overview of the overall accuracies for the whole data set and per variant for different classifiers. Classification of 14 gesture classes independent of variants resulted in an overall accuracy of up to 60 % using SVMs with the RBF kernel, which is more than 8 times higher than classification by random guessing (≈7 %). SVMs with the RBF kernel performed slightly better than the Bayesian classifiers, the SVMs with the linear kernel and the neural networks. Decision trees performed worse than the other classifiers.

Classification within each gesture variant showed that the accuracies for the rough variants (up to 62 %) were higher than for the normal variants (up to 60 %), which were higher than those for the gentle variants (up to 54 %). The exception was the Bayesian classifier. In this case the normal variants performed slightly better than the rough variants. The SVM classifiers (both kernels) performed slightly better than the Bayesian classifier and neural network. Again, decision trees performed worse than the other classifiers.

5 Discussion

In this section we will discuss the touch gesture recognition results in depth, looking into accuracy differences between subjects and between different classifiers, the interaction between gestures and the different variants and confusions between touch gestures. Also, we will reflect critically on the collection of the touch gesture data.

5.1 Classification results and touch gesture confusion

From the classification results in Table 6 it can be seen that the more complex classifiers (i.e., SVM and neural network) performed better than the simpler decision tree. However, the performance of the simpler Bayesian classifier was only slightly lower than those of the SVM and neural network. This indicates that recognition rates are reasonably robust across different classification methods.

The subject independent model generalized well for some subjects but not for others as shown by the large individual differences in accuracy for the total data set in Table 7. Differences in accuracy between subjects ranged from 44 % for the Bayesian classifiers and the decision trees to 50 % for the linear SVMs and neural networks. These individual differences make it harder to build a reliable subject independent model for touch gesture recognition. Depending on the application a trade-off can be made to build subject-dependent models which could increase accuracy at the expense of the need for training sessions. Between classifiers, results per subject differed on average 13 %. These differences were largely due to the overall lower decision tree results, per subject the accuracies for the other four classifiers differed on average 6 %. As expected, gentle gestures were considerably harder to classify which can be due to the lower pressure levels used for this gesture variant (see Table 3), resulting in a lower signal-to-noise ratio.

To gain insight into the interaction between gestures and their variants, we classified the gesture variants (i.e., 3 classes) and the combination of gestures and their variants (i.e., 42 classes) using the Bayesian classifier as a baseline due to its simplicity. Classification of the gesture variants using leave-one-subject-out cross-validation yielded accuracies ranging from 39 to 64 % ($M = 50$ %, $SD = 6$ %). Over participants the correct rate for the classification of all gestures dependent on variant ranged from 15 to 47 % ($M = 32$ %, $SD = 9$ %). Misclassification was most common between the gestures' variants which is in line with the low accuracy for the classification of the gesture variants. Confusions between gestures were similar to those found for classification independent of gesture variant. For example: 'gentle grab' was correctly classified in 36 % of the samples and was most often misclassified as 'normal grab' (24 %), 'gentle squeeze' (16 %), 'normal squeeze' (8 %) and 'rough grab' (5 %).

Misclassification was mostly due to confusions between similar touch gestures. Table 8 shows the confusion matrix for the SVM with the RFB kernel of the whole data set as this classifier yielded the best results. The five most frequently confused gesture pairs were: grab and squeeze (sum of 294 confused samples); pat and tap (280); rub and stroke (223); scratch and tickle (219); hit and slap (154). Within gesture variants the rankings of most confused pairs were similar to those of the combined variants. Also, confusions between touch gestures depicted in Table 8 largely matched the touch gesture pairs that were reported to be difficult for the participants in Sect. 3.6. However, some small differences were observed: although 'pinch vs. squeeze' was in the top five most often reported difficulties in the confusion matrix this was not one of the most frequently confused gesture pairs

Table 7 Accuracy per participant for all data for the different classifiers

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Bayesian	.42	.58	**.78**	**.70**	**.71**	.45	.58	.66	.56	**.75**	.66	.54	.60	.37	.47	.41	.65	.58	.50	.51	.56	.58	.34	.66	.41	.69	.60	.54	.55	.58	.58
Decision tree	.39	.48	.65	.63	.58	.43	.52	.55	.52	.67	.54	.47	.48	.30	.39	.29	.62	.48	.45	.41	.50	.52	.23	.56	.36	.54	.48	.42	.55	.50	.51
SVM linear	.52	.58	**.83**	**.73**	**.76**	.52	.60	.65	.58	**.78**	**.71**	.63	.59	.41	.48	.44	.58	.65	.57	.52	.56	.67	.33	.65	.44	**.70**	.62	.44	.55	.63	.63
SVM RBF	.51	.58	**.82**	**.72**	**.76**	.50	.61	.68	.56	**.80**	**.72**	.63	.62	.42	.42	.47	.65	.65	.60	.52	.56	.66	.37	.65	.44	**.72**	.65	.46	.58	.65	.65
Neural network	.48	.59	**.79**	**.74**	**.73**	.47	.63	.66	.59	**.82**	**.71**	.61	.62	.44	.40	.47	.61	.63	.57	.49	.56	.65	.32	.67	.46	**.74**	.64	.49	.54	.63	.62

Legend–accuracy: ≥*50 %*, ≥**70 %**

(sum of 104 confused samples). Conversely, 'scratch vs. tickle' was one of the five most confused gesture pairs but was not among the most often mentioned difficulties (mentioned by three participants).

Recognizing a large set of different touch gestures can reduce the classification accuracy, especially when gestures show many overlapping characteristics. Therefore, it is important to find the right balance for each application. To illustrate this trade-off we composed a subset of gestures by starting with the original 14 gestures and removing one of the gestures for each of the five most commonly confused gesture pairs, the subset consisted of nine gestures: grab, massage, pinch, poke, press, slap, stroke, tap and tickle. Classification of this gesture subset independent of variant using a Bayesian classifier with leave-one-subject-out cross-validation yielded accuracies ranging from 45 to 94% ($M = 75\ \%, SD = 12\ \%$). The performance increased by 18 % for the recognition of nine touch gestures compared to the results with fourteen touch gestures using the same classifier. However, at the cost of the ability to distinguish between more classes.

To get an indication of the most important features, the top five features for each optimized decision tree using leave-one-subject-out cross-validation were listed (i.e., the first five splits). Table 9 shows the top features ranked on frequency. While it is possible for features to appear multiple times in the top 5 with different cut-off values this was not the case for the features displayed here. Therefore, the maximum frequency for the features listed is equal to the number of cross-validation folds (=31). These five highest frequency features were among the most important features for most trained decision trees indicating that these features are reasonably robust. Mean pressure of the 7th sensor row was found to be an important feature for all trees. The 7th sensor row was positioned on the side of the mannequin arm facing away from the participant. When the hand was (partially) folded around the arm it is supposed that the fingers pressed down on this sensor area. A possible explanation for the importance of this feature is that the level of pressure in this sensor area can indicate whether the hand is folded around the arm as would be expected for gestures such as grab and squeeze.

No gender differences were observed based on basic touch gesture characteristics (see Table 5). To look for more subtle differences we classified the touch gestures based on gender using a Bayesian classifier and 10-fold cross-validation. Accuracies ranged from 75 to 78 % ($M = 76\ \%, SD = 0.01\ \%$), which is similar to the baseline accuracy when classifying every sample as 'male' ($24/31 \approx 77\ \%$). Based on our findings we have no reason to assume that gender differences play a significant role in touch gesture classification. However, it should be noted that our sample size does not allow us to rule out possible differences.

Table 8 Confusion matrix of leave-one-subject-out cross-validation using SVMs with RBF kernel for all data (overall accuracy = 60 %)

	Actual class													
	Grab	Hit	Massage	Pat	Pinch	Poke	Press	Rub	Scratch	Slap	Squeeze	Stroke	Tap	Tickle
Predicted class														
Grab	**397**	0	17	1	11	1	31	4	4	0	*177*	2	0	0
Hit	1	**317**	0	45	1	15	1	0	0	*77*	3	1	45	0
Massage	4	0	**386**	2	1	1	0	*63*	26	1	14	11	1	26
Pat	8	*58*	1	*268*	1	2	4	1	22	*59*	0	12	*149*	18
Pinch	3	4	6	1	**398**	27	25	8	8	0	*66*	1	6	3
Poke	1	27	0	11	*68*	**438**	50	0	2	4	3	1	40	5
Press	19	4	0	7	30	25	**374**	17	1	7	23	6	8	2
Rub	0	0	*78*	2	1	0	8	*239*	*56*	0	2	*98*	0	40
Scratch	6	0	7	5	0	0	2	50	*274*	0	0	12	0	*92*
Slap	0	*77*	0	*70*	2	0	0	2	1	**358**	0	14	44	1
Squeeze	*117*	0	15	0	38	0	50	1	2	0	*268*	1	0	0
Stroke	0	1	28	8	1	0	2	*125*	34	3	0	**383**	4	15
Tap	0	*68*	0	*131*	6	46	11	2	1	48	0	2	*248*	16
Tickle	2	2	19	6	0	3	0	45	*127*	1	1	12	13	**339**
Sum	558	558	557	557	558	558	558	557	558	558	557	556	558	557

Legend–classification of touch gesture captures into a class: ≥*10 %*, ≥ **50** %

Table 9 Features that were most frequently ranked within the top 5 for decision tree classification using 31-fold cross-validation

Feature (no.)	Frequency
Mean pressure of the 7th sensor row (10)	31
Summed traveled in column direction (27)	30
Average spatial peak distance to center off mass (39)	30
Overall mean pressure difference between frames (43)	30
Highest pressure contact area (21)	27

Accuracies reported in this work were higher than the accuracy of 53 % that was previously reported for the CoST data set using a Bayesian classifier [23]. This indicated that the additional features and the use of more complex classification methods with hyperparameter optimization have improved the accuracy. Results reported in this paper fall within the range of 26–61 % accuracy that was reported for a data challenge using the CoST data set [3,12,18,36]. Our results are comparable to those reported in the work of Gaus et al. and Ta et al. who reported accuracies up to 59 and 61 %, respectively using random forest [12,36]. However it should be noted that the data challenge contained a subset of CoST (i.e., gentle and normal variants) and that the train and test data division was different from the leave-one-subject-out cross-validation results reported in this paper [24]. For an overview of the data challenge and the challenge protocol the reader is refereed to [24].

5.2 Considerations regarding the data collection

The instructions during the data collection were given in English to include non-native Dutch speakers. However, this could have resulted in translation discrepancies between the English language and the participants' native language. Silvera-Tawil et al. [32] gave the example of the back-translation of the word 'pat' from Spanish to English which can be either translated to 'pat' or 'tap'. Based on observations in a pilot test we opted to include visual examples of the different touch gestures rather than providing participants with the definitions in Table 2 to reduce the language barrier. The use of visual examples instead of giving text-based definitions could however have reduced the interpersonal differences as participants might have tried to mimic the examples.

To minimize the influence on the participants' natural touch behavior we opted for not restricting the time taken for every touch gesture. Also, there were no constraints on the number of instances of a touch gesture that could be part of a single capture. A consequence of this decision is that a single tap and three taps are both treated as a single touch gesture. This raises the question whether a single tap has a different meaning than three consecutive taps. Furthermore, as features were calculated from the segmented data, segmentation has an influence on features that cover gesture duration (e.g. gesture duration in frames).

The sensor data was labeled according to the instructions (i.e., if the participant was instructed to perform a 'gentle

grab', the corresponding sensor data was labeled as such). During segmentation some touch gesture captures were filtered out based on minimal change in gesture intensity, successfully removing skipped touch gestures. However, this procedure does not control for all possible mistakes, which makes it probable that the dataset contains incorrect labels. Manual annotation of the video recordings could help filter out mistakes such as cases where a touch gesture was performed that was different from the one that was instructed.

The inclusion of touch gesture variants seemed to have increased the diversity of the ways in which the touch gestures were performed. Descriptive statistics confirmed that participants used pressure to distinguish between the gesture variants, using less than normal pressure for the gentle variants and more than normal pressure for the rough variants. The definitions of the gentle variant and the rough variant given by the participants also indicated that the amount of pressure is an important way to distinguish between the two for example by the use of the keywords 'soft' and 'hard'. Although speed is also used to differentiate between gentle and rough as indicated by the use of the keywords 'slow' and 'fast', respectively. The downside is that the reliance on pressure to distinguish between both the gestures and the different variants of the same gestures has probably increased the difficulty of the touch gesture recognition. Notably, in the definitions from Table 2 the use of words such as 'forcible', 'gently' and 'firmly' again point to the importance of force/pressure and also temporal component are mentioned (e.g. 'quickly', 'repeatedly'). As these characteristics seem to be inherent to some of the touch gestures, one may argue that a roughly performed pat, which should generally be 'gently and quickly', would resemble more of a slap, which should generally be 'quickly and sharply'. The claim that some gestures do not lend themselves as easily for variants is further supported by the self reports of the participants.

6 Conclusion and future work

To study automatic touch recognition we collected CoST, a data set containing 7805 gesture captures of 14 different touch gestures each performed in three variants: gentle, normal and rough. The data showed similarities between gestures and large differences in the way these gestures were performed which was increased by the inclusion of the different gesture variants. From the different classifiers that were compared, the best results were obtained using SVMs with the RBF kernel while the decision tree yielded the worst performance. Classification of the 14 touch gestures independent of the gesture's variant yielded an average accuracy of up to 60 %. The subject independent model generalized well for some individuals but not for others. Gentle gesture variants proved to be harder to classify than the normal and rough variants. Misclassification was most common between touch gestures with similar characteristics.

Further research into highly discriminating features using feature selection or dimension reduction can be beneficial for applications that require (on-board) real-time touch classification in which computational power is costly. Also, at this moment it is unclear what the minimum requirements are regarding touch recognition performance in order to have a meaningful touch interaction.

Furthermore, to behave socially intelligently, an artifact such as a robot, should not only be able to sense and recognize touch gestures but should also be able to interpret those touch gestures in order to respond in a socially appropriate manner as illustrate in Fig. 1. As most studies, including the work presented here, focus on parts of the interaction, future work should tie together the full interaction cycle.

In the current setup we have studied touch recognition in a controlled lab setting with specific instructions, lacking social context which could help to recognize touch gestures and their inferred meaning. According to Hertenstein et al. the complexity of the tactile system allows for the same touch gesture to have different meanings as touch can also vary in its intensity, velocity, abruptness, temperature, location and duration [15]. The meaning of touch is also dependent on factors such as concurrent verbal and nonverbal behavior, the type of interpersonal relationship and the situation in which the touch takes place [16,21].

As touch is only one of the modalities that plays a role in social interaction, social signals from other modalities can provide valuable information as well. Currently we are working on a study on the interpretation of touch behavior within a social context. Human interaction with a robot pet companion is observed by looking at touch behavior as well as other social behaviors such as speech and gaze and the role of these behaviors in touch interaction. To further close the interaction loop we are also looking into appropriate responses.

Acknowledgments This publication was supported by the Dutch national program COMMIT.

References

1. Altun K, MacLean KE (2015) Recognizing affect in human touch of a robot. Pattern Recognit Lett 66:31–40
2. Bailenson JN, Yee N, Brave S, Merget D, Koslow D (2007) Virtual interpersonal touch: expressing and recognizing emotions through haptic devices. Human Comput Interact 22(3):325–353

3. Balli Altuglu T, Altun K (2015) Recognizing touch gestures for social human-robot interaction. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), (Seattle, WA), pp 407–413
4. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. Trans Intell Syst Technol 2:27:1–27:27
5. Chang J, MacLean K, Yohanan S (2010) Gesture recognition in the haptic creature. In: Proceedings of the International Conference EuroHaptics, Amsterdam, The Netherlands, pp 385–391
6. Cooney MD, Becker-Asano C, Kanda T, Alissandrakis A, Ishiguro H (2010) Full-body gesture recognition using inertial sensors for playful interaction with small humanoid robot. In: Proceedings of International Conference on Intelligent Robots and Systems (IROS), Taipei, Taiwan, pp 2276–2282
7. Cooney MD, Nishio S, Ishiguro H (2012) Recognizing affection for a touch-based interaction with a humanoid robot. In: Proceedings of the International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal, pp 1420–1427
8. Dahiya RS, Metta G, Valle M, Sandini G (2010) Tactile sensing—from humans to humanoids. Trans Robot 26(1):1–20
9. Flagg A, MacLean KE (2013) Affective touch gesture recognition for a furry zoomorphic machine. In: Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction (TEI), Barcelona, Spain, pp 25–32
10. Flagg A, Tam D, MacLean KE, Flagg R (2012) Conductive fur sensing for a gesture-aware furry robot. In: Proceedings of the Haptics Symposium (HAPTICS), Vancouver, Canada, pp 99–104
11. Gallace A, Spence C (2010) The science of interpersonal touch: an overview. Neurosci Biobehav Rev 34(2):246–259
12. Gaus YFA, Olugbade T, Jan A, Qin R, Liu J, Zhang F et al (2015) Social touch gesture recognition using random forest and boosting on distinct feature sets. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), Seattle, WA, pp 399–406
13. Haans A, IJsselsteijn W (2006) Mediated social touch: a review of current research and future directions. Virtual Real 9(2–3):149–159
14. Hertenstein MJ, Verkamp JM, Kerestes AM, Holmes RM (2006) The communicative functions of touch in humans, nonhuman primates, and rats: a review and synthesis of the empirical research. Genet Soc Gen Psychol Monogr 132(1):5–94
15. Hertenstein MJ, Holmes R, McCullough M, Keltner D (2009) The communication of emotion via touch. Emotion 9(4):566–573
16. Heslin R, Nguyen TD, Nguyen ML (1983) Meaning of touch: the case of touch from a stranger or same sex person. Nonverbal Behav 7(3):147–157
17. Hsu CW, Chang CC, Lin CJ, et al (2003) A practical guide to support vector classification. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. Accessed 17 May 2016
18. Hughes D, Farrow N, Profita H, Correll N (2015) Detecting and identifying tactile gestures using deep autoencoders, geometric moments and gesture level features. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), Seattle, WA, pp 415–422
19. Jeong S, Logan D, Goodwin M, Graca S, O'Connell B, Goodenough H et al (2015) A social robot to mitigate stress, anxiety, and pain in hospital pediatric care. In: Proceedings of the International Conference on Human–Robot Interaction (HRI) Extended Abstracts, Portland, OR, pp 103–104
20. Ji Z, Amirabdollahian F, Polani D, Dautenhahn K (2011) Histogram based classification of tactile patterns on periodically distributed skin sensors for a humanoid robot. In: Proceedings of the International Symposium on Robot and Human Interactive Communication (ROMAN), Atlanta, GA, pp 433–440
21. Jones SE, Yarbrough AE (1985) A naturalistic study of the meanings of touch. Commun Monogr 52(1):19–56
22. Jung MM (2014) Towards social touch intelligence: developing a robust system for automatic touch recognition. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), Istanbul, Turkey, pp 344–348
23. Jung MM, Poppe R, Poel M, Heylen DKJ (2014) Touching the void—introducing CoST: corpus of social touch. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), Istanbul, Turkey, pp 120–127
24. Jung MM, Cang XL, Poel M, MacLean KE (2015) Touch challenge '15: recognizing social touch gestures. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), Seattle, WA, pp 387–390
25. Jung MM, Poel M, Poppe R, Heylen DKJ (2016) Corpus of social touch (CoST). University of Twente, Enschede. doi:10.4121/uuid:5ef62345-3b3e-479c-8e1d-c922748c9b29
26. Kim YM, Koo SY, Lim JG, Kwon DS (2010) A robust online touch pattern recognition for dynamic human–robot interaction. Trans Consum Electron 56(3):1979–1987
27. Knight H, Toscano R, Stiehl WD, Chang A, Wang Y, Breazeal C (2009) Real-time social touch gesture recognition for sensate robots. In: Proceedings of the International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, pp 3715–3720
28. Kotranza A, Lok B, Pugh CM, Lind DS (2009) Virtual humans that touch back: enhancing nonverbal communication with virtual humans through bidirectional touch. In: Proceedings of the Virtual Reality Conference, Lafayette, LA, pp 175–178
29. Nakajima K, Itoh Y, Hayashi Y, Ikeda K, Fujita K, Onoye T (2013) Emoballoon: A balloon-shaped interface recognizing social touch interactions. In: Proceedings of Advances in Computer Entertainment (ACE), Boekelo, The Netherlands, pp 182–197
30. Naya F, Yamato J, Shinozawa K (1999) Recognizing human touching behaviors using a haptic interface for a pet-robot. In: Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC), Tokyo, Japan, vol 2, pp 1030–1034
31. Silvera-Tawil D, Rye D, Velonaki M (2011) Touch modality interpretation for an eit-based sensitive skin. In: Proceedings of the International Conference on Robotics and Automation (ICRA), Shanghai, China, pp 3770–3776
32. Silvera-Tawil D, Rye D, Velonaki M (2012) Interpretation of the modality of touch on an artificial arm covered with an EIT-based sensitive skin. Robot Res 31(13):1627–1641
33. Silvera-Tawil D, Rye D, Velonaki M (2014) Interpretation of social touch on an artificial arm covered with an EIT-based sensitive skin. Int J Soc Robot 6(4):489–505
34. Silvera-Tawil D, Rye D, Velonaki M (2015) Artificial skin and tactile sensing for socially interactive robots: a review. Robot Auton Syst 63:230–243
35. Stiehl WD, Lieberman J, Breazeal C, Basel L, Lalla L, Wolf M (2005) Design of a therapeutic robotic companion for relational, affective touch. In: Proceedings of International Workshop on Robot and Human Interactive Communication (ROMAN), Nashville, TN, pp 408–415
36. Ta VC, Johal W, Portaz M, Castelli E, Vaufreydaz D (2015) The Grenoble system for the social touch challenge at ICMI 2015. In: Proceedings of the International Conference on Multimodal Interaction (ICMI), Seattle, WA, pp 391–398
37. Wada K, Shibata T (2007) Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. Trans Robot 23(5):972–980
38. van Wingerden S, Uebbing TJ, Jung MM, Poel M (2014) A neural network based approach to social touch classification. In: Proceedings of Workshop on Emotion Representation and Modelling in Human–Computer-Interaction-Systems (ERM4HCI), Istanbul, Turkey, pp 7–12
39. Yohanan S, MacLean KE (2012) The role of affective touch in human–robot interaction: human intent and expectations in touching the haptic creature. Soc Robot 4(2):163–180