

Touching the Void – Introducing CoST: Corpus of Social Touch

Merel M. Jung¹, Ronald Poppe², Mannes Poel¹, Dirk K. J. Heylen¹

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

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

m.m.jung@utwente.nl, r.w.poppe@uu.nl, m.poel@utwente.nl, d.k.j.heylen@utwente.nl

ABSTRACT

Touch behavior is of great importance during social interaction. To transfer the tactile modality from interpersonal interaction to other areas such as Human-Robot Interaction (HRI) and remote communication automatic recognition of social touch is necessary. This paper introduces CoST: Corpus of Social Touch, a collection containing 7805 instances of 14 different social touch gestures. The gestures were performed in three variations: gentle, normal and rough, on a sensor grid wrapped around a mannequin arm. Recognition of the rough variations of these 14 gesture classes using Bayesian classifiers and Support Vector Machines (SVMs) resulted in an overall accuracy of 54% and 53%, respectively. Furthermore, this paper provides more insight into the challenges of automatic recognition of social touch gestures, including which gestures can be recognized more easily and which are more difficult to recognize.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: Haptic I/O; I.5.2 [PATTERN RECOGNITION]: Design Methodology—*Classifier design and evaluation*

General Terms

Measurement, Performance

Keywords

Social touch; Touch corpus; Touch gesture recognition

1. INTRODUCTION

Touch behavior is one of the important non-verbal forms of social interaction as are visual cues such as facial expressions and body gestures [16]. In interpersonal interaction, touch is important for establishing and maintaining social interaction [5]. Also, touch is used to generate and communicate both positive and negative emotions as well as to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions.acm.org.

ICMI'14, November 12–16, 2014, Istanbul, Turkey.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-2885-2/14/11 ...\$15.00.

<http://dx.doi.org/10.1145/2663204.2663242>

express intimacy, power, and status [5, 7]. The positive effects of touch on well-being, such as stress reduction, are extensively described in the literature (for a review see [4]). Furthermore, it is known that tactile interaction may affect compliance and liking [5, 7]. Touch behavior is seen in many different forms of social interaction: a handshake as a greeting, a high-five to celebrate a joint accomplishment, a tap on the shoulder to gain someone's attention, a comforting hug from a friend, or holding hands with a romantic partner.

The human sense of touch consists of physiological inputs from various receptors: receptors in the skin register pressure, pain and temperature while receptors in the muscles, tendons and joints register body motion [5]. However, just equipping a robot or interface with touch sensors is not enough. Automatic recognition of social touch is necessary to transfer the tactile modality from interpersonal interaction to other areas such as Human-Robot Interaction (HRI) and remote communication [5, 6]. Providing tactile social intelligence to robots and virtual agents can open up opportunities for various applications. If a robot can understand social touch behavior, the robot can respond accordingly – resulting in richer and more natural interaction. One of these applications is robot therapy in which robots are used to comfort people in stressful environments such as children in hospitals and elderly people in nursing homes [15].

Some promising attempts have been made to recognize different touch gestures for specific applications. However, recognition rates depend on the degree of similarity between the gestures. In order to engage in tactile interaction with a robot or an interface, there is a need for reliable recognition of a wide range of social touch gestures. A Robust touch gesture recognition system should be real-time and generalizable across users. Because there is a need for touch datasets, we have recorded a corpus of social touch gestures to characterize various touch gestures and work towards reliable recognition. The main contributions of this paper are a Corpus of Social Touch (CoST) and the insights from a first exploration into the recognition of these touch gestures.

The remainder of the paper is organized as follows: the next section discusses related work on recognition of social touch; Section 3 presents CoST; classification results of the gestures from a subset of the data are presented and discussed in Sections 4 and 5, respectively; The paper concludes in Section 6.

2. RELATED WORK

Previous attempts to sense, classify and interpret social touch behavior include the development and deployment of

several animal and humanoid robots, as well as the use of touch for other applications such as remote communication. One approach is to focus on building an artificial skin for robots that simulate the human somatosensory system using touch sensors [3]. To allow for physical interaction between humans and robots, Naya et al. intended to cover a robot with sheets of pressure sensitive ink. Five different touch gestures: ‘pat’, ‘scratch’, ‘slap’, ‘stroke’ and ‘tickle’, were performed on a single 44×44 gridded sensor sheet [12]. The absolute pressure values and total touch surface were found to be discriminative features for the touch gestures ‘pat’, ‘scratch’ and ‘slap’ using the k-nearest neighbor method. ‘Stroke’ and ‘tickle’ could be distinguished using temporal differences in pressure and touch surface using Fisher’s linear discriminant method. Combining the results of two classifications using different classifiers and features resulted in an overall accuracy of 87% between subjects. Silvera-Tawil and colleagues developed an artificial skin to enable tactile HRI based on the principle of Electrical Impedance Tomography to measure pressure [13]. Six touch gestures: ‘pat’, ‘push’, ‘scratch’, ‘slap’, ‘stroke’ and ‘tap’ were performed on a flat surface covered with the artificial skin. Classification of the gestures using the LogitBoost algorithm resulted in overall accuracies of 91% within a single subject and 74% between multiple subjects. The same six touch gestures were performed by the same participants on the back of a single person who acted as a human classifier resulting in an average performance of 86%. In a follow-up study, Silvera-Tawil et al. compared human touch recognition on the arm with automatic recognition on a full-sized mannequin arm covered with the artificial skin [14]. Nine gestures: ‘no-touch’, ‘pat’, ‘pull’, ‘push’, ‘scratch’, ‘slap’, ‘stroke’, ‘squeeze’, and ‘tap’ were classified by human receivers and the LogitBoost algorithm. Human recognition ($M = 90\%$) was higher than automatic classification using leave-one-subject-out cross-validation ($M = 71\%$).

Another approach is to focus on the embodiment of a specific robot which can be covered with sensors. *The Huggable* is a robotic companion in the form of a teddy bear [15]. In an initial study nine touch gestures: ‘contact’, ‘pat’, ‘pet’, ‘poke’, ‘rub’, ‘scratch’, ‘slap’, ‘squeeze’ and ‘tickle’ were performed on an arm of the robot equipped with temperature sensors (thermistors), proximity sensors (electric field sensors) and pressure sensors (Quantum Tunneling Composite sensors) [15]. Using a neural network to classify 199 gesture instances showed that some touch gestures such as ‘rub’ and ‘squeeze’ could be recognized from the sensor data, however, ‘slap’ could not be recognized. *The Sensate Bear* platform was developed to explore the feasibility of real-time classification of social touch before integration with *The Huggable* teddy bear [10]. To sense social touch, the body of the robot bear was covered with capacitive sensors to register proximity and contact area as well as to distinguish between humans and objects. Real-time classification of four gestures showed that ‘foot-rub’ and ‘head-pat’ could be accurately recognized while ‘hug’ and ‘tickle’ were more problematic. ‘Hugs’ were difficult to sense because sensing through clothing was not possible while ‘tickle’ was difficult to recognize because of the large variation in body locations. *Haptic Creature* is a robot that resembles a small lap animal which can sense and react to touch input [1, 17]. A first study into the recognition of four touch gestures – ‘pat’, ‘poke’, ‘slap’ and ‘stroke’ – performed by a single participant was based on

the data of the force sensing resistors which were attached all over the body of the *Haptic Creature* [1]. Accuracies ranged from 11% to 77% depending on the sensor density of the part of the body that was touched. Cooney et al. used a mock-up of a humanoid robot to study affectionate touch behavior towards the mock-up [2]. Two types of touch sensors using photo-interrupters were incorporated in the upper body of the humanoid mock-up: one for detecting perpendicular movement (i.e. towards and from the body surface) such as patting and the other for detecting lateral movement (i.e. parallel to the body surface) such as rubbing. Both computer vision (using Microsoft Kinect) and the touch sensors were used to recognize twenty different touch interactions such as ‘pat head’, ‘rub back’ and ‘shake hand’. Also, the performance of two classifiers: Support Vector Machines (SVMs) with radial basis function kernel and the k-nearest neighbor method were compared using leave-one-subject-out cross-validation. Overall accuracies showed that the SVM classifiers consistently outperformed the k-nearest neighbor method. Recognition for the SVMs of touch gestures using only vision (78%) performed better than using only touch (72%) while the combination of both modalities yielded the best performance (91%).

Aside from tactile HRI, the recognition of social touch can also be used for other types of interfaces which can, for example, enable remote communication. *Emoballoon* is a balloon interface for social touch which contains a barometric pressure sensor and a microphone [11]. Seven different touch gestures: ‘grasp’, ‘hug’, ‘press’, ‘punch’, ‘rub’, ‘slap’ and ‘no-touch’ were classified using SVMs with the radial basis function kernel resulting in an overall accuracy of 75% between participants and 84% within participants. In a study of Huisman et al. on communicative mediated touch, participants wore a tactile sleeve containing a pressure sensitive input layer and an output layer consisting of vibrotactile motors [9]. Subjects received six different touch gestures: ‘hit’, ‘poke’, ‘press’, ‘rub’, ‘squeeze’ and ‘stroke’ on the vibrotactile display of the sleeve which they had to imitate using the sleeve’s input layer. Comparison of the duration and contact area used for the prerecorded gestures and the imitated gestures showed that people had difficulty with the precise replication of the touch duration and touch surface. Participants were not directly asked to classify the gestures, however, human classification of the gestures based on video recordings of the imitated gestures showed that the received gesture often differed from the imitated gesture especially for the gestures ‘rub’ and ‘stroke’.

In summary, development of an artificial skin to provide future robots with a sense of touch is beneficial but brings extra design requirements such as flexibility and stretchability to cover curved surfaces and moving joints [13, 14]. In the short term, the use of a fully embodied robot covered with sensors has the advantage of providing information about body location which can be used to recognize touch [2, 10]. However, this can cause problems ensuring adequate sensor density on all body parts [1]. Furthermore, Silvera-Tawil et al. showed that comparable accuracies can be achieved using partial embodiment in the form of an arm covered with artificial skin [13, 14]. Automatic classification on several sets of touch gestures, ranging from 4 to 20 different gestures, performed on various robots, robot skins and interfaces yielded mixed results. For example, recognition varied per gesture [10, 15] and location [1]. Human classification

of touch gestures outperformed automatic classification [13, 14]. However, human gesture recognition of mediated touch was found to be more difficult [9].

In order to work towards more robust gesture recognition, further research on the characteristics of touch gestures is needed. One central problem is that there is no gesture data set available for research and benchmarking. This work focuses on improving touch gesture recognition instead of on embodiment of a robot/ interface by collecting a data set containing a relatively large set of touch gestures.

3. COST: CORPUS OF SOCIAL TOUCH

To address the need for social touch datasets, we recorded a corpus of social touch gestures. The data set is publicly available¹. CoST consists of sensor data from 31 participants performing 3 variations (normal, gentle and rough) of 14 different touch gestures. Each gesture was performed 6 times on a sensor grid wrapped around a mannequin arm. The arm was chosen as the contact surface because it is one of the least invasive body areas to be touched [8] and presumably a neutral body location to touch others. The data from the pressure sensor consists of a pressure value (i.e. intensity) per channel (i.e. location) at 135 fps (i.e. temporal resolution).

3.1 Touch gestures

The touch gestures (see Table 1) were taken from the touch dictionary of [17]. The list of gestures was adapted to suit interaction with an artificial human arm. Touch gestures involving physical movement of the arm itself, for example lift, push and swing, were omitted because the movement of the mannequin arm could not be sensed by the pressure sensor grid. In the instructions of the gesture to perform, the participants were shown the name of the gesture and not the definition from [17]. Instead, they were shown an example video before the start of the data collection in which each gesture was demonstrated. During the data collection, 14 different touch gestures were performed 6 times in 3 variations, resulting in 252 gesture instances per participant. The order of gestures was pseudo-randomized into three blocks using the following rule: each instruction was given two times per block but the same instruction was not given twice in succession. A single fixed list of gestures was constructed using this criterion. This list and the reversed order of the list were used as instructions in a counterbalanced design. After each touch gesture, the participant had to press a key to see the next gesture. The keystrokes were used for segmentation afterwards. Figure 1 shows the evolution over time of the summed pressure for a gesture instance of each class. Figure 2 and Figure 3 show the evolution over time of the summed pressure for the gesture ‘rough grab’ performed by a single participant and multiple different participants, respectively. These examples illustrate the variation in duration and intensity of a gesture, within participants and between participants.

3.2 Setup

For the sensing of the gestures, an 8×8 pressure sensor grid (from *plug-and-wear*, www.plugandwear.com) was connected to a Teensy 3.0 USB Development Board (by *PJRC*, www.pjrc.com). The sensor is made of textile consisting of

¹Data available on request, contact m.m.jung@utwente.nl

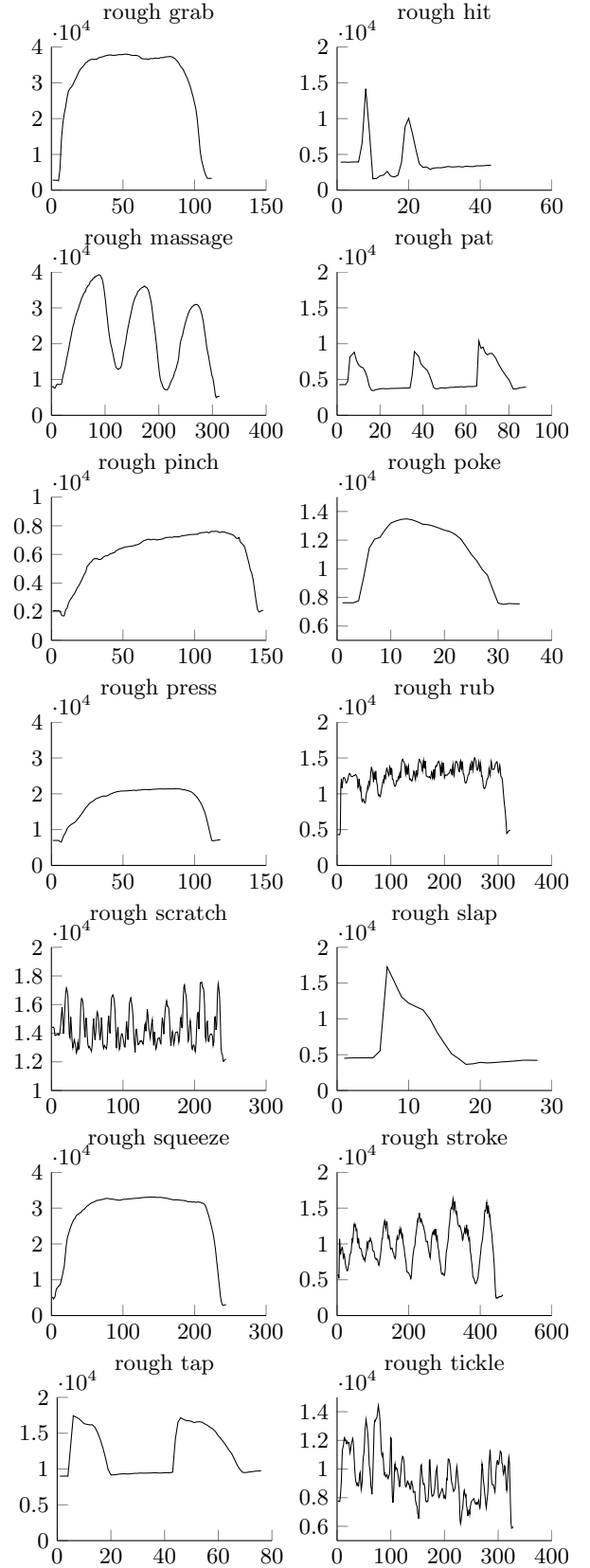


Figure 1: Summed pressure (y-axis) over time (x-axis) for a gesture instance of each class.

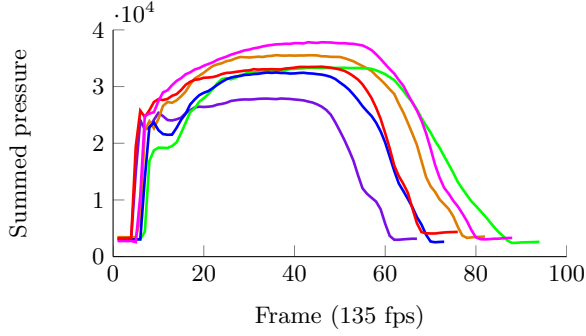


Figure 2: Summed pressure per frame of a ‘rough grab’ performed six times by a single participant.

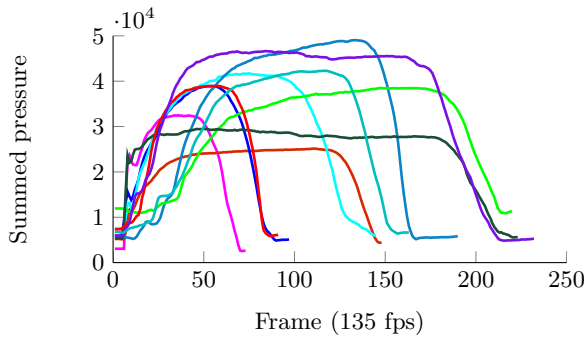


Figure 3: Summed pressure per frame of a ‘rough grab’ performed by multiple different participants.

Table 1: Touch dictionary adapted from [17].

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.

five layers. The two outer layers are protective layers made of felt. Each outer layer is attached to a layer containing eight strips of conductive fabric separated by non-conductive strips. Between the two conductive layers is the middle layer which comprises a sheet of piezoresistive material. The conductive layers are positioned orthogonally so that they form an 8 by 8 matrix. One of the conductive layers is attached to the power supply while the other is attached to the A/D converter of the Teensy board. The sensors’ detectable pressure ranges from 1.8×10^{-3} to > 0.1 MPa at an ambient temperature of 25°C . After A/D conversion, the pressure values of the 64 channels range from 0 to 1023 (i.e., 10 bits). Sensor data was sampled at 135 Hz. 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 Figure 4). The mannequin arm was fastened to the table to prevent it from slipping. Video recordings were made during the data collection as verification of the sensor data and the instructions given. The instructions for which gesture to perform had been scripted and were displayed to the participants on a PC monitor.

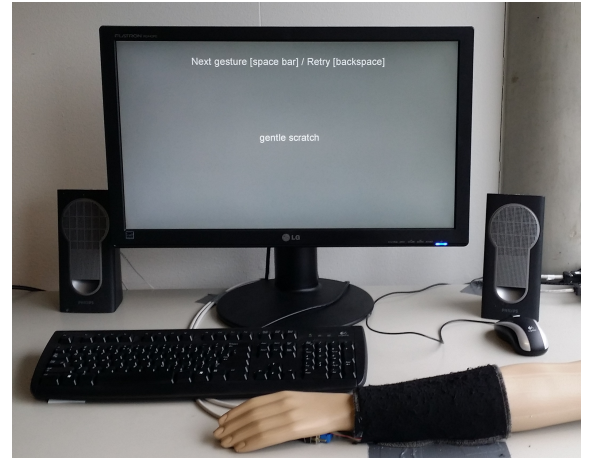


Figure 4: Set-up showing the pressure sensor (the black fabric) wrapped around the mannequin arm and the computer monitor displaying the instructions.

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. The age of the participants ranged from 21 to 62 years ($M = 34$, $SD = 12$) and 29 were right-handed.

3.4 Preprocessing

The raw data was checked and segmented before feature extraction. Each subset of variations consists of 14 gestures \times 6 repetitions \times 31 people = 2604 gestures in total. Coarse segmentation was based on the key strokes of the participants marking the end of a gesture. As segmentation between keystrokes still contained many frames from before and after the gesture, the data was further segmented. Removing these additional frames is especially important to reduce noise in the calculation of features which use a time

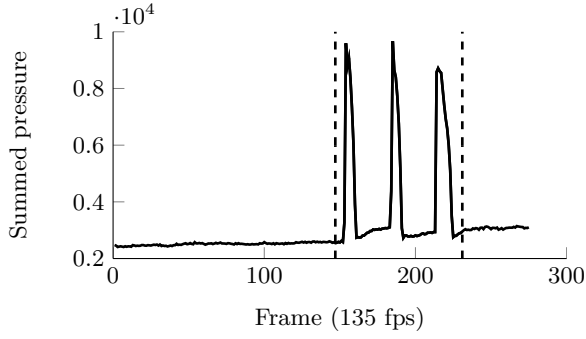


Figure 5: Segmentation of a ‘rough pat’ based on pressure difference indicated by the dashed lines.

component, such as averaging over frames and the total amount of frames of a gesture instance. Further segmentation was based on the change in the gesture’s intensity, that is, the summed pressure over all 64 channels. Parameters were optimized by visual inspection and kept constant for the whole data set. See Figure 5 for an example of the segmentation of a ‘rough pat’.

The automatic segmentation was inspected visually to ensure that all gestures were captured. Inspection of the segmented data showed that six gesture instances could not be automatically segmented because there were only small differences in pressure during the gesture. The video recordings revealed that an instance of a ‘rough stroke’ was performed too fast to be distinguishable from the sensors’ noise. The other five gesture instances were accidentally skipped. One other notable gesture instance of a ‘normal squeeze’ was of much longer duration (over a minute) than all other instances. The video footage showed that instead of a single squeeze, all 14 different touch gestures were practiced again on the sensor grid, while the data recording was already started. All seven gesture instances were removed from the dataset. See Table 2 for total of gesture instances per variation after preprocessing.

Table 2: Gesture instances per variation and in total of the CoST dataset after preprocessing.

	Variation			Total
	Gentle	Normal	Rough	
Recorded	2604	2604	2604	7812
Data loss	1× massage, 1× pat, 1× stroke	1× tickle, 1× squeeze	1× rub, 1× stroke	7
Data	2601	2602	2602	7805

4. RECOGNITION OF TOUCH GESTURES

For the first exploration into the recognition of social touch gestures from the CoST dataset, the rough touch gestures were used because of the favorable signal-to-noise ratio. The rough gesture subset consisted of 2602 gesture instances.

4.1 Feature extraction

From the sensor data, 28 features were extracted. Features are numbered between brackets.

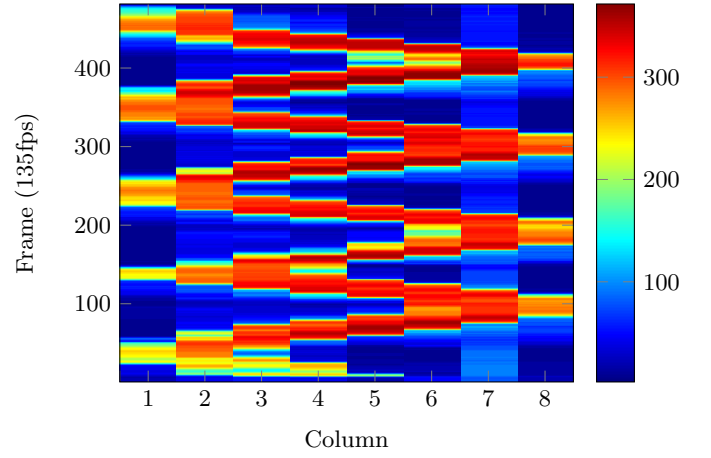


Figure 6: A ‘rough stroke’. Mean pressure per sensor grid column over time visualizes the displacement in the opposite direction (i.e. the rows).

Mean pressure was calculated by the mean pressure of all channels averaged over time (1).

Maximum pressure is the maximum channel value of the gesture (2).

Pressure variability indicates the differences in pressure applied during the gesture. The variability over time was calculated by the mean absolute pressure differences summed over all channels (3).

Mean pressure per column was calculated over time, resulting in eight features (4-11).

Mean pressure per row was calculated over time, resulting in eight features (12-19).

Contact area was calculated per frame as the percentage of sensor area (i.e. the number of channels divided by the total numbers of channels) that exceed 50% of the maximum pressure. Two features were calculated: the mean of the contact area over time (20) and the contact area of the frame with the maximum overall pressure (i.e. the highest summed pressure over all channels) (21).

Peak count is the number of positive crossings of the threshold. Two ways to calculate the threshold were used, resulting in two features: one threshold is defined as 50% of the frame with the maximum summed pressure (22), the other is defined as the mean of the summed pressure over time (23).

Displacement indicates whether the area of contact is static during a gesture or whether the hand moves across the contact area (i.e. dynamic). Figure 6 shows an example of a dynamic gesture (a ‘rough stroke’). The position of the center of mass is used to calculate the movement on the contact surface in both the x-axis and the y-axis. Four features were calculated, both the mean over time and the summed absolute difference of the center of mass on the x-axis (24-25) and the y-axis (26-27).

Duration is the time used to make contact with the surface to perform the gesture which is measured in frames at 135 fps (28).

To visualize similarities within and between gestures, a normalized distance matrix between the summed pressure over time of the gestures was calculated using the dynamic time warping algorithm to indicate the difficulty of the clas-

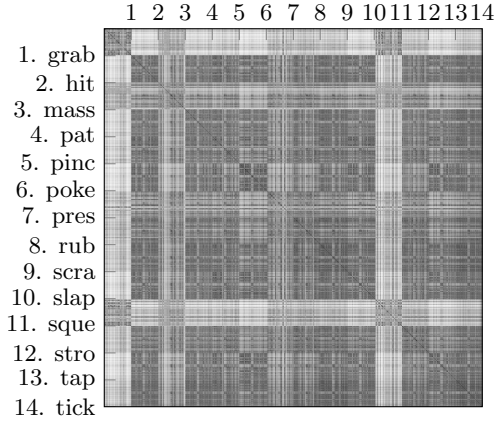


Figure 7: Normalized distance matrix of the summed pressure over time for all rough gesture instances. Darker areas indicate smaller differences.

sification problem (see Figure 7). Darker areas indicate smaller differences between two gesture instances. See Figure 8 for the mean pressure of all channels averaged over time (feature 1) and the duration (feature 28) plotted per gesture class. It can be seen from the figure that there is a lot of overlap between gesture classes and a large spread within classes.

4.2 Classification

The features from Section 4.1 were used for classification using MATLAB[®] (release 2013b). First, Gaussian Bayesian classifiers were used as baseline performance for gesture recognition. Second, the more complex SVM classifiers were used for comparison. The results of the classification were evaluated using leave-one-subject-out cross-validation. The baseline of classifying a sample into the correct class is $1/14 \approx 7\%$.

4.2.1 Bayesian classifiers

The mean and covariance for each class were calculated from the training data. The parameters for the multivariate normal distribution were used to calculate the posterior probability of the test sample belonging to the given class. Samples were assigned to the class with the maximum posterior probability. The summed results of the 31-fold cross-validation are displayed in a confusion matrix in Table 3. Between participants, the correct rate over all classes ranged from 24% to 75% ($M = 54\%$, $SD = 12\%$).

4.2.2 Support Vector Machine classifiers

We treated the classification of touch gestures as a multi-class problem using the one-vs.-all approach using the linear kernel with the default parameter of $C = 1$. Using this approach, models were trained for every gesture versus all other gestures. A test sample was classified for all models as belonging to the gesture class or the other class. There were 3 possible scenarios: (1) the test sample was classified as belonging to only one gesture class, in that case the test sample was assigned to that respective class, (2) the test sample was classified as belonging to multiple gesture classes, in that case the test sample was assigned to the class with the maximum distance to the hyperplane, (3) the test

Table 3: Confusion matrix of the Bayesian classifiers [Overall accuracy = 54%]. *Legend* – classification of gesture instances into a class: $\geq 10\%$, $\geq 50\%$.

		Actual class													
		grab	hit	mass	pat	pinc	poke	pres	rub	scra	slap	sque	stro	tap	tick
Predicted class	1	121	0	1	0	0	0	18	2	0	0	73	0	0	0
	2	0	116	0	26	0	6	1	0	0	53	0	0	31	0
	3	7	0	127	0	2	1	0	18	12	0	7	1	0	8
	4	0	17	3	43	0	5	0	4	4	18	0	14	26	5
	5	2	2	12	4	125	18	15	5	5	2	26	5	1	3
	6	0	7	0	7	16	131	15	0	1	2	1	0	18	3
	7	8	3	0	6	25	8	115	11	5	2	14	4	3	0
	8	0	2	17	4	0	0	2	76	16	0	0	33	4	10
	9	0	1	3	4	0	1	1	18	96	0	0	8	0	40
	10	0	25	0	20	0	0	0	1	93	0	1	17	0	0
	11	46	1	7	2	17	1	15	0	1	1	65	2	1	0
	12	0	1	9	4	0	1	2	37	11	1	0	109	3	3
	13	0	11	0	61	1	13	2	0	0	12	0	2	78	5
	14	2	0	7	5	0	1	0	14	34	2	0	6	4	109
sum		186	186	186	186	186	186	186	185	186	186	186	185	186	186

Table 4: Confusion matrix of the SVM classifiers [Overall accuracy = 53%]. *Legend* – classification of gesture instances into a class: $\geq 10\%$, $\geq 50\%$.

		Actual class													
		grab	hit	mass	pat	pinc	poke	pres	rub	scra	slap	sque	stro	tap	tick
Predicted class	1	123	0	5	0	1	0	14	1	0	0	67	0	0	0
	2	0	123	0	22	0	4	0	0	0	64	0	1	23	0
	3	3	0	126	0	4	1	2	18	8	0	10	1	0	6
	4	0	27	0	80	1	4	9	4	6	18	0	4	39	5
	5	9	5	3	2	129	54	18	0	1	0	33	0	2	0
	6	0	4	1	0	13	86	17	0	0	1	0	0	11	1
	7	1	2	1	6	9	5	68	5	2	3	2	1	3	1
	8	0	0	22	2	2	0	26	62	12	0	2	25	3	8
	9	4	0	6	3	3	6	0	23	102	3	1	5	2	46
	10	1	14	0	14	0	1	2	2	0	82	1	5	14	0
	11	42	0	3	0	9	0	13	1	0	0	65	2	0	0
	12	1	0	7	6	11	1	11	55	14	2	3	135	4	3
	13	0	11	0	45	3	20	6	2	1	13	2	3	81	4
	14	2	0	12	6	1	4	0	12	40	0	0	3	4	112
sum		186	186	186	186	186	186	186	185	186	186	186	185	186	186

sample was never classified as belonging to the gesture class, in that case the test sample was assigned to the class with the minimum distance to the hyperplane. The summed results of the 31-fold cross-validation are displayed in a confusion matrix in Table 4. Between participants the correct rate over all classes ranged from 32% to 75% ($M = 53\%$, $SD = 11\%$).

4.2.3 Comparison of classifiers

Overall accuracy of both classifiers was more than 7 times higher than classification by random guessing ($1/14$). However, there were large difference between gestures and participants. The correct rate per gesture for both the Bayesian classifiers and the SVMs are listed in Table 5. The mean correct rates and the standard deviations of the two classifiers were almost equal ($M = 54\%$, $SD = 14\%$ vs. $M = 53\%$, $SD = 14\%$). Differences between the two classifiers were calculated by subtracting the correct rate per gesture of the Bayesian classifiers from the correct rates of the SVMs, the differences ranged from -20% to 25% ($M = 1\%$, $SD = 12\%$). The most notable differences were seen for the gestures ‘poke’ and ‘press’ which were classified much more accurately ($\geq 20\%$) using Bayesian classifiers while ‘pat’ was classified much more accurately ($\geq 20\%$) using SVMs.

The most frequent confusion was between the following gestures: ‘grab’ and ‘squeeze’; ‘hit’, ‘pat’, ‘slap’ and ‘tap’; ‘rub’ and ‘stroke’; ‘scratch’ and ‘tickle’. Confusion between ‘grab’ and ‘squeeze’ can be explained by the similarity in contact area, use of the whole hand, and the duration of the gesture. Furthermore, it can be argued that ‘grab’ is

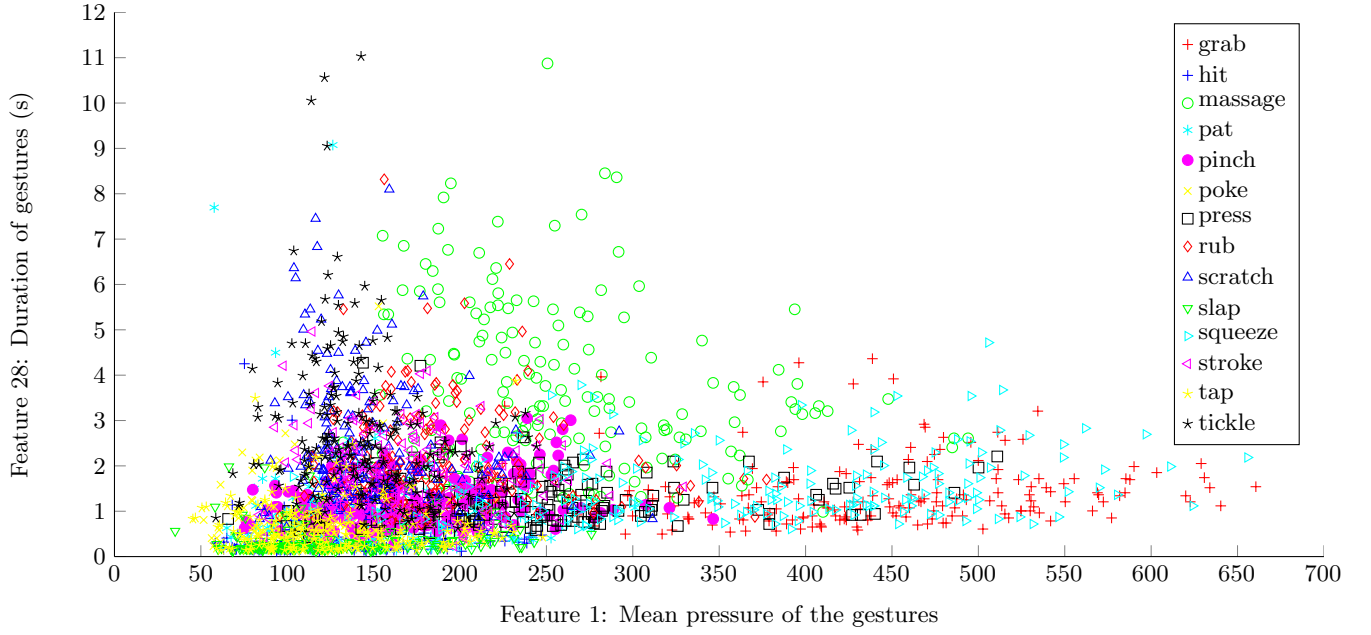


Figure 8: Feature 1 and feature 28 plotted per gesture class. Best viewed in color.

Table 5: Comparison of accuracy per gesture for both classifiers. Legend – accuracy: $\geq 50\%$, $\geq 70\%$.

Gesture	grab	hit	mass	pat	pinc	poke	pres	rub	scra	slap	sque	stro	tap	tick
Bayes	.65	.62	.68	.23	.67	.70	.62	.41	.52	.50	.35	.59	.42	.59
SVM	.66	.66	.68	.43	.69	.46	.37	.34	.55	.44	.35	.73	.44	.60
Difference	-.01	-.04	.00	-.20	-.02	.24	.25	.07	-.03	.06	.00	-.14	-.02	-.01

part of the ‘squeeze gesture’. ‘Hit’, ‘pat’, ‘slap’ and ‘tap’ show similarities in duration, contact area and possibility of having been repeated several times during one gesture instance. In contrast, differences in pressure are expected between ‘hit’ and ‘slap’ compared to ‘pat’ and ‘tap’. Both ‘rub’ and ‘stroke’ are prolonged gestures in which a back and forth movement is expected, however higher pressure levels are expected for ‘rub’. ‘Scratch’ and ‘tickle’ are both characterized by a frequent change of direction and long duration. However, for ‘tickle’, pressure levels are expected to be lower and more variability in direction is expected instead of the back and forth movement of ‘scratch’. The inclusion of touch gesture variants could have increased the difficulty of differentiating between gestures because pressure is one of the main characteristics on which the classification was based. Instructions to perform the touch gestures in gentle, normal and rough variations, could have encouraged subjects to use pressure levels to differentiate between gesture variants (e.g. ‘gentle pat’ vs. ‘rough pat’) rather than between gesture classes (e.g. ‘pat’ vs. ‘slap’). However, in natural settings force differences can be expected between subjects based on personal characteristics such as physical strength but also within subjects based on the social context.

The correct rate per participant of the leave-one-subject-out cross-validation for both the Bayesian classifiers and the multi-class SVMs are listed in Table 6. The mean correct rates and standard deviations of the two classifiers were almost equal ($M = 54\%$, $SD = 12\%$ vs. $M = 53\%$, $SD = 11\%$). Differences between the two classifiers were calcu-

lated by subtracting the correct rate per participant of the Bayesian classifiers from the correct rates of the SVMs, the differences ranged from -8% to 14% ($M = 1\%$, $SD = 6\%$).

5. DISCUSSION

Classification of 14 gesture classes resulted in a maximum overall accuracy of 54% which is less than reported in related studies which reported mostly accuracies $> 70\%$. However, direct comparison of gesture recognition between studies based on reported accuracies is difficult because of differences in gesture sets, sensors, and classification protocols. Some reported accuracies were the result of a best-case scenario intended to be a proof of concept [1, 12]. While other studies focused on the location of the touch rather than the gesture, such as distinguishing between ‘head-pat’ and ‘foot-rub’ [10] or ‘handshake’ and ‘back-pat’ [2]. Although it could be argued that some gestures will be more suitable for particular body locations. Also, within subjects results were generally better than between subjects [11, 13]. However, some studies have used a subset of the gesture set in this paper which makes comparison possible. Recognition of a subset of the gestures was evaluated with Bayesian classifiers following the same procedure as used for the full gesture set described in Section 4.2.1. Using the gesture set of [1]: ‘pat’, ‘poke’, ‘slap’ and ‘stroke’, resulted in an overall accuracy of 79% which is comparable with the overall accuracy of 77% reported by Chang et al. on the area with the largest sensor density. Recognition of the gesture set of [12]: ‘pat’, ‘scratch’, ‘slap’, ‘stroke’ and ‘tickle’, resulted in an overall accuracy of 69%. This result is less than the overall accuracy of 87% reported by Naya et al., however these results were achieved by combining the maximum accuracies per class of two classifiers. Recognition of the gesture set of [9]: ‘hit’, ‘poke’, ‘press’, ‘rub’, ‘squeeze’ and ‘stroke’ resulted in an overall accuracy of 77% however, direct com-

Table 6: Comparison of accuracy per participant for both classifiers. Legend – colors: $\geq 50\%$, $\geq 70\%$.

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
Bayes	.52	.69	.71	.56	.75	.48	.58	.54	.57	.75	.64	.50	.61	.35	.49	.36	.49	.58	.50	.38	.57	.62	.24	.68	.45	.63	.54	.44	.51	.52	.49
SVM	.52	.56	.68	.64	.75	.50	.58	.49	.64	.75	.65	.51	.48	.35	.46	.44	.51	.51	.51	.36	.43	.62	.32	.65	.44	.62	.55	.32	.55	.46	.50
Difference	.00	.13	.03	-.08	.00	-.02	.00	.05	-.07	.00	-.01	-.01	.13	.00	.03	-.08	-.02	.07	-.01	.02	.14	.00	-.08	.03	.01	.01	-.01	.12	-.04	.06	-.01

parison was not possible because accuracies per class were not provided.

Preliminary classification results of the normal and gentle gestures indicate similar performance to the rough gesture variants. Bayesian classifiers (normal: $M = 56\%$, $SD = 11\%$; gentle: $M = 47\%$, $SD = 11\%$) outperformed the SVMs (normal: $M = 48\%$, $SD = 10\%$; gentle: $M = 41\%$, $SD = 10\%$). The gentle gestures were more difficult to distinguish which was expected due to the lower signal-to-noise ratio.

6. CONCLUSION

In order to work towards reliable recognition of social touch gestures this paper introduces a Corpus of Social Touch (CoST) to fill the void of available touch data. This paper describes CoST, a data set consisting of 14 different touch gestures as well a first exploration into the automatic recognition of the social touch gestures of CoST. Classification of the rough variations showed that 14 gesture classes could be classified with an overall accuracy of 54% and 53% using Bayesian classifiers and SVMs respectively. Results were more than 7 times higher than chance ($1/14 \approx 7\%$) but less than results reported in related literature. The difference can partly be explained by the use of different sets of touch gestures, within person classification and the use of different types of sensors. Classification of a subset of gesture classes used in other studies resulted in overall accuracies which were more similar to the accuracies reported in the literature. Future work will include further analysis of the CoST dataset as well as optimization of the gesture recognition.

7. ACKNOWLEDGMENTS

This publication was supported by the Dutch national program COMMIT.

8. REFERENCES

- [1] J. Chang, K. MacLean, and S. Yohanan. Gesture recognition in the haptic creature. In *Proceedings of the International Conference EuroHaptics*, (Amsterdam, The Netherlands), pages 385–391. 2010.
- [2] M. D. Cooney, S. Nishio, and H. Ishiguro. 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), pages 1420–1427, 2012.
- [3] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini. Tactile sensing — from humans to humanoids. *Transactions on Robotics*, 26(1):1–20, 2010.
- [4] T. Field. Touch for socioemotional and physical well-being: A review. *Developmental Review*, 30(4):367–383, 2010.
- [5] A. Gallace and C. Spence. The science of interpersonal touch: an overview. *Neuroscience & Biobehavioral Reviews*, 34(2):246–259, 2010.
- [6] A. Haans and W. IJsselstein. Mediated social touch: a review of current research and future directions. *Virtual Reality*, 9(2-3):149–159, 2006.
- [7] M. J. Hertenstein, J. M. Verkamp, A. M. Kerestes, and R. M. Holmes. The communicative functions of touch in humans, nonhuman primates, and rats: a review and synthesis of the empirical research. *Genetic, Social, and General Psychology Monographs*, 132(1):5–94, 2006.
- [8] R. Heslin, T. D. Nguyen, and M. L. Nguyen. Meaning of touch: The case of touch from a stranger or same sex person. *Journal of Nonverbal Behavior*, 7(3):147–157, 1983.
- [9] G. Huisman, A. Darriba Frederiks, B. Van Dijk, D. Heylen, and B. Kröse. The tasst: Tactile sleeve for social touch. In *Proceedings World Haptics Conference (WHC)*, (Daejeon, Korea), pages 211–216, 2013.
- [10] H. Knight, R. Toscano, W. D. Stiehl, A. Chang, Y. Wang, and C. Breazeal. Real-time social touch gesture recognition for sensate robots. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, (St. Louis, MO), pages 3715–3720, 2009.
- [11] K. Nakajima, Y. Itoh, Y. Hayashi, K. Ikeda, K. Fujita, and T. Onoye. Emoballoon a balloon-shaped interface recognizing social touch interactions. In *Proceedings of Advances in Computer Entertainment (ACE)*, (Boekelo, The Netherlands), pages 182–197. 2013.
- [12] F. Naya, J. Yamato, and K. Shinozawa. 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), volume 2, pages 1030–1034, 1999.
- [13] D. Silvera-Tawil, D. Rye, and M. Velonaki. Touch modality interpretation for an eit-based sensitive skin. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, (Shanghai, China), pages 3770–3776, 2011.
- [14] D. Silvera-Tawil, D. Rye, and M. Velonaki. Interpretation of the modality of touch on an artificial arm covered with an eit-based sensitive skin. *The International Journal of Robotics Research*, 31(13):1627–1641, 2012.
- [15] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf. Design of a therapeutic robotic companion for relational, affective touch. In *Proceedings of International Workshop on Robot and Human Interactive Communication (ROMAN)*, (Nashville, TN), pages 408–415, 2005.
- [16] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland. Social signals, their function, and automatic analysis: a survey. In *Proceedings of the international conference on Multimodal interfaces (ICMI)*, (Chania, Crete, Greece), pages 61–68, 2008.
- [17] S. Yohanan and K. E. MacLean. The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature. *International Journal of Social Robotics*, 4(2):163–180, 2012.