Social Touch Gesture Recognition using Random Forest and Boosting on Distinct Feature Sets

Yona Falinie A. Gaus
Brunel University London
yonafalinie.abdgaus@brunel.ac.uk

Temitayo Olugbade
UCL Interaction Centre
UCL
temitayo.olugbade.13@ucl.ac.uk

Asim Jan
Brunel University London
asim.jan@brunel.ac.uk

Rui Qin
Brunel University London
rui.qin@brunel.ac.uk

Jingxin Liu
Brunel University London
jingxin.liu@brunel.ac.uk

Fan Zhang
Brunel University London
fan.zhang@brunel.ac.uk

Hongying Meng[*]
Brunel University London
hongying.meng@brunel.ac.uk

Nadia Berthouze
University College London
nadia.berthouze@ucl.ac.uk

ABSTRACT

Touch is a primary nonverbal communication channel used to communicate emotions or other social messages. A variety of social touch exists including hugging, rubbing and punching. Despite its importance, this channel is still very little explored in the affective computing field, as much more focus has been placed on visual and aural channels. In this paper, we investigate the possibility to automatically discriminate between different social touch types. We propose five distinct feature sets for describing touch behaviours captured by a grid of pressure sensors. These features are then combined together by using the Random Forest and Boosting methods for categorizing the touch gesture type. The proposed methods were evaluated on both the HAART (7 gesture types over different surfaces) and the CoST (14 gesture types over the same surface) datasets made available by the Social Touch Gesture Challenge 2015. Well above chance level performances were achieved with a 67% accuracy for the HAART and 59% for the CoST testing datasets respectively.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: User Interfaces; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Touch*

Keywords

Social Touch, Touch Gesture Recognition, Touch Features

1. INTRODUCTION

[*]Corresponding author.

Touch is an important channel for affective communication and experience, as such interactive embodied technology should be endowed with the capabilities to interpret it. Despite this growing interest in various disciplines, and clear opportunities for affective-touch aware technology, this affective modality is still very little explored in the affective computing community where the focus has been on other affective modalities ([44] [43] [24] [2] [33] [34]). The Social Touch Gesture Challenge 2015 is timely attempting to stimulate interest and lead to the development of social touch automatic classification systems as the community has done for the other modalities. This paper aims to contribute to it by proposing ways to describe touch gesture types (e.g., tapping, caressing) and methods to automatically categorize them.

Humans communicate emotions and other social messages through touch. We explore and appreciate objects through touch. The emerging understanding of the physiological and neural mechanisms at the basis of affective touch [9],[40] has led to a growing interest in investigating how we communicate or express affect through touch. Studies on multimodal communication, have shown that touch amplify the intensity of an emotion conveyed by the face and through the voice [25]. Burgoon [5] showed that touch itself can convey more complex social feelings such as trust and affection. Hertenstein et al. [18] [17] contributed to this body of work by showing that through touch we do not only communicate the intensity of an emotion, but also its valence and that we are able to discriminate between at least nine discrete emotions from touch behaviour alone.

In the field of Human-Computer Interaction, touch has been explored as a way to build intimacy and facilitate distance relationships. For example, Park et al. [31] have shown that, when given an haptic channel, people develop haptic codes to communicate at distance with their partners. In human-robot interaction, social touch has been shown to increase bond and trust in various therapeutically applications ([35] [36] [20]). In the commercial and design settings, [1] has shown that we use different types of touches not only to un-

derstand the characteristics of an object (textile in this case) but also the affective sensations it provokes in us. This is further supported reported in [32], where the authors has shown that, during creative design sessions, designers use affective touch to stimulate their creative processes. They also discuss how designers observe the hands of their customers touching a product to understand what they feel. In more general terms, [21] and [3] discuss the importance and the possibilities that affective touch offers to evaluate the user experience in human-computer interaction contexts.

Initial work towards building systems for the automatic detection of affective touch are summarised in Gao et al. [13]. Through their study, the authors also shows that the kinematics of strokes during a touch-based computer games on smartphone allow the detection of how a player feels with very high performances. However, each of these works focus on one touch type at the time (e.g., tapping, stroking). Discriminating between different touch types is however very important, as shown by Hertenstein [17]. In addition, for affective and social touch to be ubiquitous, it is important that the ability to classify it is independent of the type and shape of a surface being touched. This has become very critical as embedding touch sensors in any type of mobile surface is becoming achievable at a relatively low cost. Hence, it is mandatory that this channel is investigated beyond the flat surface of touch-based displays. The Social Touch Gesture Challenge 2015 answers to these needs by providing two touch behaviour datasets: the 'Human-Animal Affective Robot Touch'(HAART) dataset [10] and the 'Corpus of Social Touch' (CoST) [23]. These datasets provide a wide range of type behaviour on surfaces of different shapes and consistencies embedded with pressure sensors. Our paper responds to this challenge by proposing a way to describe the types of touch gestures tracked by the sensors and by proposing an ensemble of machine learning methods to discriminate between the different types of touch behaviour.

The rest of the paper is organized as follows. Section 2 provides an overview of related works. Afterwards, Section 3 illustrates the methodology of our proposed approach including the description of feature sets and classification methods. Section 4 presents the experimental results on the challenge datasets. The possible influencing factors for different performance indicators are discussed in Section 5 together with future directions.

2. RELATED WORKS

Human-Robot interaction is possibly one of the more active areas where researchers have been investigating how to create touch sensing capabilities. A typical approach is to cover a robot with an artificial skin that simulates human touch sensory system [8] [30] [42] [37] [39]. An indirect way to give robots the capability to sense touch is instead to embed them with sensors that measure temperature, proximity and pressure [41] [19] [29] [46] [7] . Works using these two types of approaches are discussed below.

The earliest work on artificial skin is in 1999, where a robot was covered with gridded pressure-sensitive conductive ink sheets . Five different touch gestures ('pat', 'scratch', 'slap', 'stroke' and 'tickle') were performed on the sensor sheet. Absolute pressure values and temporal difference between

pressure values were used as features to discriminate between these different types of touch. Using K-Nearest Neighbor classification method, it was possible to discriminate between 'slap', 'pat' and 'scratch' by simply using contact area and absolute pressure values. For more similar gestures of 'stroke' and 'tickle', both temporal difference and contact area were instead needed and the Fisher Linear Discriminant classification method [30] was used. These approaches led to an overall accuracy of 60%.

Silvera-Tawil et al [42] developed further the concept of artificial skin by using the principle of electrical impedance tomography (EIT). This innovation enabled to cover larger areas of a robot and allowed to extract information such as location, duration, displacement and intensity of touch. 'pat', 'push', 'stroke', 'slap' and 'tap' gestures were investigated and the LogitBoost algorithm was used for classification purpose. The overall accuracy reached on average 91% for individual participant, but decreased to 74% when modelling multiple participants together due to the variation of touch patterns between individual participants. Individual variability was also observed in Gao et al. [13] in terms of stroke kinematics during touch-based games. In a follow-up study, Silvera-Tawil et al [37] successfully implemented an artificial skin to cover the arm of a full-size mannequin. Increasing the number of touch gesture by adding 'pull', 'slap' and 'squeeze', they compared the classification accuracy obtained with the LogitBoost algorithm (71%) with human recognition (90%). Further experiments were shown in [38] by interpreting social touch on same artificial skin. Up to 90% accuracy were achieved using Logitboost algorithm for both emotions and social messages, similar with human recognition.

The second approach simulates touch perception capabilities by using sensors. This approach has been extensively used in social robotics: Shibata's baby seal Paro [35] [36], SONY's dog, Aibo [12] [11], Stiehl's teddy bear the Huggable [41], Saldien and Goris's elephant-like creature Probo [15] [14], and the latest one is a furry artificial lap-pet called Haptic Creature [6] [46] [10]. Apart from robot-animal like form, other types of interface such as Emoballon [29], TaSSt project [19] can enable remote communication. We cover below some of the more significant differentiating factors in relation to these robots or interfaces.

The Huggable [41] is a robotic companion with the form of a teddy bear, equipped with a combination of temperature, electric field, and QTC force sensors intended to facilitate affective haptic communication between two people. Nine touch gestures ('contact', 'pat', 'pet', 'poke', 'rub', 'scratch', 'slap', 'squeeze' and 'tickle') were performed on an arm of the robot, and features such as change in direction of motion, average sensor value and the number of sensors active among others, were extracted from the sensors. By using supervised neural network as classification method, it was found that only 'slap' gesture were not well classified. This was due to delays in the sensor reading during 'slap' gestures as these gestures are generally very fast.

Probo [15] [14] is an elephant-like social robot equipped with a large variety of sensors. Probo can recognize whether it is being hugged, scratched, or hurt. The robot is used to ad-
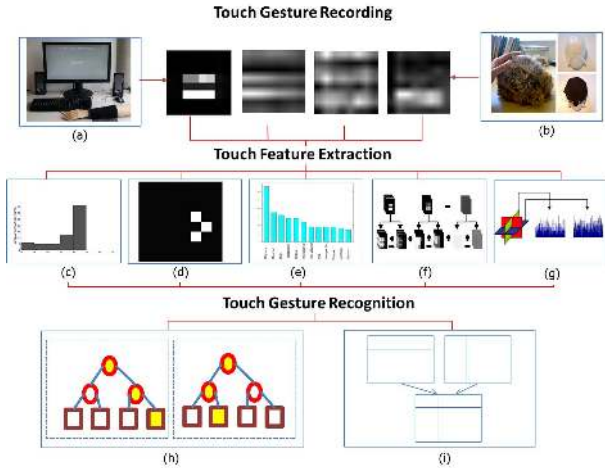
Figure 1: Overview of the proposed social touch gesture recognition system (a) CoST (b) HAART (c) SD (d) BMH (e) MSD (f) SMMHH (g) LBPTOP (h) Random Forest (i) Boosting.

dress anxiety in hospitalized children. Similarly, Paro [35] [36], the famous robot seal, responds to being touched, differentiates between being stroked and hit, and tries to amend its own behaviour accordingly, repeating actions that have been rewarded with stroking, and avoiding actions that have led to being hit. The Haptic Creature, developed by Yohanan and his colleagues [45] [46], is a furry lap-sized social robot that communicates with the world through touch. By expressing itself through ear stiffness modulations, breathing rate, and purring patterns, the Haptic Creature aims to create a comforting experience through touch-based interactions by sensing and responding to human touch. The first experiments with the Haptic Creature investigated the recognition of four touch gesture types - 'pat', 'stroke', 'slap', 'poke'- by attaching force sensitive resistors all over its body [6]. These four gestures were differentiated with an average of 77% accuracy, depending on the sensor density of where it was touched.

Taking one step further, Yohanan and Maclean [46] examined how humans express their feelings for the Haptic Creature by means of touch gestures. Based on these reports, Yohanan and MacLean compiled a touch dictionary, also called the Yohanan's dictionary. Their findings inspired [10] and [23] to create two datasets containing labelled pressure sensor data of social hand touch gestures: HAART and CoST. The aim of these datasets is to challenge the creation of automatic affective touch classification systems by providing a large variety of touch behaviour types and kinematics as well as surfaces of different shapes and consistencies.

3. METHODOLOGY
This section introduces the two datasets of touch types used in this study, the set of features proposed to describe the touch instances and the methods to classify them. The overall system is shown in Fig 1.

3.1 Touch Gesture Capture

Table 1: HAART dataset: 7 touch gestures with different substrates and covers.

Substrate	Cover	Gesture
CURVE	Fur	constant, no touch,
FOAM	Long	pat, rub, tickle,
NONE	None	scratch, stroke

Table 2: COST dataset: 14 touch gestures.

1	2	3	4	5	6	7
grab	hit	massage	pat	pinch	poke	press
8	9	10	11	12	13	14
rub	scratch	slap	stroke	squeeze	tap	tickle

This research is based on the two datasets provided by the Social Touch challenge 2015: HAART and CoST.

3.1.1 HAART dataset
HAART [10] contains 7 touch gestures, performed by 10 subjects on an 8-by-8 grid of pressure sensors, where each sensor is approximately 1 square inch wide. The touch behaviour was recorded at 54 Hz with 64 sensors of data cell per frame. The sensors return integer values ranging between 0 and 972. Each subject performed each gesture under different combinations of substrates and different cover conditions with the sensor lying between the substrate and the surface cover. These are summarized in Table 1.

3.1.2 CoST dataset
CoST [23] contains 14 types of touch gestures (see Table 2). The touch gestures were performed by 31 subjects on a 8-by-8 grid of pressure sensors wrapped around a mannequin arm. The gestures were recorded with a frame rate of 135 Hz with 64 channels per frame. The values obtained from each channel are integers ranging between 0 and 1023. Each participant performed each gesture 6 times with 3 variations: gentle, normal and rough. More details on the data can be found in [23].

3.2 Touch Feature Extraction
The original recording of the touch sensors provided by the Social Touch Gesture Challenge 2015 needed to be transformed into high-level features, to gain more meaningful descriptions of the touch behaviour. The aim is to develop general features that are as independent as possible of the surface considered and of the kinematics of the gesture type. Parameters tuning is used to optimize their discriminative power to the different recording settings used for the two dataset but not between surfaces within the HAART dataset. Five sets of features are proposed.

3.2.1 Statistical distribution (SD) of pressure surface
We aimed to identify how much surface area (i.e., the number of sensor cells) is either not touched, touched (i.e., we are sure the area is touched and it is not noise) or touched with high pressure. The values used to capture these three conditions are represented in Table 3. For the no-touch and high-pressure conditions, we computed two features for each dataset: one using values computed at gesture level and one using values computed at frame level. For the touched condition, the median pressure value was used as a threshold
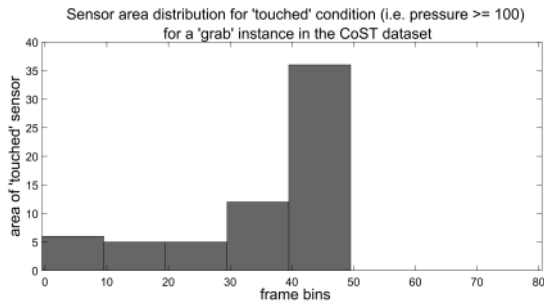
Figure 2: An example of a sensor area distribution feature for the 'touched' condition for a 'grab' instance using the MATLAB function histcounts and the relevant parameters listed in Table 4.

indicating that the surface had been touched. In the case of the CoST, an analysis of the data, revealed that a pressure value of 100 was also a good threshold for capturing when the surface had been touched.

For each gesture instance and for each pressure value/range (Table 3), we computed the underlying distribution of the number of sensor cells per frames having such value/range. The MATLAB 'histcounts' was used to this purpose. In order to normalize the length of these features within a dataset (given that gestures had different frame numbers), appropriate bin widths and edges (see Table 4) were set to fit all instances. These parameters were estimated using the same Matlab function. An example of resulting feature (represented by a histogram) is provided in Fig 2.

Table 4: Parameters computed for the MATLAB function 'histcount'

	CoST		HAART	
	bin width	bin edges leading, trailing	bin width	bin edges leading, trailing
$h_{notouch}$	1	-0.5, 11.5	5	-0.5, 55.5
$h_{touched}$	10	-0.5, 80.5	5	-0.5, 65.5
$h_{highpre.}$	10	-0.5, 100.5	5	-0.5, 55.5

3.2.2 Binary Motion History (BMH)

This set of features were used to capture the 'shape' of a gesture by computing a binary motion history (BMH) for each gesture instance. For each instance, an 8 by 8 matrix representing the sensor grid is created. A cell of the matrix is set to 1 if throughout the gesture instance, the corresponding sensor cell was touched at least once with high pressure (CoST: pressure >= (maximum − 100); HAART: pressure >= (maximum − 5)), 0 otherwise. Fig. 3 shows examples of BMH for CoST gestures: 'grab', 'pinch', 'poke', 'rub', 'tickle', and 'pat' instances (in order from left to right). These examples illustrate how the BMH features provide an intuitive description of the different shapes from the types of the various gestures.

3.2.3 Motion statistical distribution (MSD)

Twelve statistic features are extracted from each sensor sequence $F = (f_1, f_2, \cdots, f_L)$ including minimum, maximum,
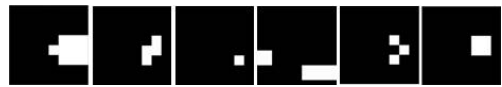

Figure 3: CoST dataset: Binary Motion History images: (from Left): Grab, Pinch, Poke, Rub, Tickle, and Pat.

mean, first quartile, median, third quartile, area, total variation, interquartile range, variance, skewness and kurtosis on the changing information for each sensor. 'L' is the number of the frame. 'area' is the sum of all values of each sensor in all frames; 'total variation' is computed as the sum of changes in values for each sensor in all frames. The number of frames for HAART dataset is fixed as 432 frames, but the number of frames for CoST dataset varies between instances. The extracted features are based on the whole gesture movement of each sensor independently of the number of frames.

3.2.4 Spatial Multi-scale Motion History Histogram (SMMHH) on touch dynamic

In order to study the touch dynamics of the recording, the data was also treated as a video. On each frame, 8-by-8 sensors were treated as 8-by-8 pixels. Motion History Histogram (MHH) is a method used for temporal-based motion analysis, detecting pixel movements from a temporal domain and summarizing them as patterns in a spatial domain. This method was first developed for human action recognition by Meng et al.[27] to discriminate between several motion gestures such as walking, boxing, jumping etc. (for further details see [26] and [28]). A variation of the MHH method, called 1-D Motion History Histogram, was suggested by Jan et al. to apply this concept onto a 1-dimensional feature space [22], which looks for binary patterns within the feature vector sequence.

We propose a new variation to the original Motion History Histogram (MHH) algorithm, with two concepts that have been applied to alter the MHH feature. The first concept is the Multi-Scale Motion History Histogram (MMHH) achieved by introducing a new dimension to MHH and by skipping a fixed range of 1 to 5 frames during the motion detection process. By comparing frame n with $n + 2 : 6$ instead of $n + 1$, the motion is greatly amplified by the visual differences across the short time range, resulting in capturing motion at different speeds. This gives a more dynamic feature than the original MHH feature, providing a different viewpoint for each gesture as shown in Fig. 4.

The second concept uses the average spatial-pooling on the reshaped data. This allows to extract the motion information from the different scales in a similar way of the convolution neural networks of deep learning.

For the touch challenge, we applied Spatial Pooling three times, producing 4 videos (including the original one) of: 8x8, 4x4, 2x2 and 1x1 sensors resolutions per gesture. On each of the videos we then extract the Multi-Scale MHH feature with appropriate thresholds. All the features produced are then resized and concatenated together making a feature

Table 3: Pressure values per datset

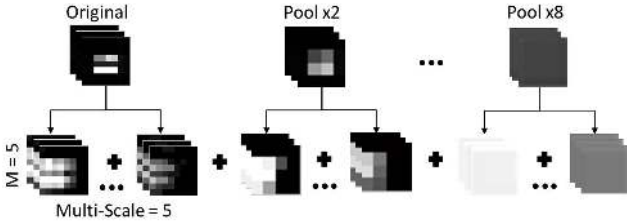|	CoST		HAART	
	At frame level	**At gesture instance level**	**At frame level**	**At gesture instance level**
no touch	= min. pressure	= min. pressure	= min. pressure	= min. pressure
touched	–	>= median pressure; >= 100	–	>= median pressure
high pressure	>= (max. pressure - 100)	>= (max. pressure - 100)	= max. pressure	= max. pressure

Figure 4: Overview of the Spatial Multi-Scale Motion History Histogram (SMMHH) on Touch Dynamic. Average pooling is applied on the video data and then the Multi-Scale Motion History Histogram is used to capture the motion data across different scales of time.

vector with the dimension of:

$$\sum_{i=1}^{4}(S \times M \times P_i) = 2125 \qquad (1)$$

where i refers to each of the Spatially pooled videos, P_i is the total number of pixels in a frame for each video i, $S = 5$ is the size of the Multi-Scale dimension, which has been set from 2:6 and $M = 5$ is the MHH pattern sequence size [27].

3.2.5 *Local Binary Pattern on Three Orthogonal Place (LBPTOP) on touch dynamic*

Similar to the SMMHH feature where touch recording data was treated as a video, the popupar video dynamic feature LBPTOP [47] is extracted.

3.3 Touch Gesture Recognition

One of the goals of this challenge was to build automatic social touch gesture recognition systems. Due to multiple distinct feature sets, ensemble learning methods were chosen in order to combine these features in an optimised way. Within ensemble learning, Random Forest [4] and Boosting methods have been selected for the classification process.

3.3.1 *Random Forest*

The Random Forest (RF) algorithm was selected given its popularity and its generally good performance in solving classification problems. The main idea behind this algorithm is to form a forest by training and combining different kinds of decision trees, and the final classification result is decided through a voting over these trees. The method combines the 'bagging' idea and the random selection of features. Detailed information can be found in [4].

3.3.2 *Boosting Algorithm*

The Boosting algorithm is another ensemble learning method that can convert a weak learner into a strong one. It gives different weights to different features and combines all distinct features together in an optimized way. The boosting implementation algorithm here is called Stage-wise Additive Modelling [16]. The Boosting algorithm was used here is similar to that of Adaboost, but the difference is in how the error rate is computed. The Boosting algorithm gives to erroneously classified data more weight.

4. EXPERIMENTAL RESULTS

The experiments adhered to the Social Touch Gesture Challenge 2015 protocol using the training and testing subsets provided. The models were firstly trained on the training subset and evaluated using a 10 fold cross-validation process. This process was also used for optimizing the parameter set for the feature extraction and for building the models. Once the optimization was completed, the parameters were fixed and the models were tested on the testing subsets by the challenge's organizer (i.e., the labels of the testing datasets were unknown to us). Both HAART and CoST datasets were used.

The Random Forest and the Boosting learning methods were used for the gesture recognition in the testing. For RF, 1000 trees were used. In the Boosting implementation, the Random Forest was used as the weak leaners. 500 iterations were used in all experiments. Both are implemented using Weka software. Accuracy was used as performance measure, that is the percentage of testing samples correctly classified with respect to the true class.

4.1 HAART

In the training subset, only the data from 7 subjects out of 10 subjects were given for a total of 578 gesture instances.

Table 5 shows the experimental results for both the individual feature sets and for the combination of all the feature sets. In the training sets, the average results of 10 fold cross validation were being used as an evaluation purpose to take advantage of all available data. When considering only individual features, MSD, SMMHH and LBPTOP obtained higher recognition rates. However, the combination of all the features led to significant improvement for the training subset, giving 76.7% accuracy on RF and 77.7% for Boosting. Taking an advantage on combination of features, the networks were trained on all sets of training data and testing on separate sets of testing data. The effect of modelling all training data together decreases accuracy to 66.5% for RF and 64.5% for Boosting. However, despite the drops in recognition accuracy in testing data, it is still indicate consistent results across both classifiers.

The confusion matrix in Table 6 and Table 7 respectively, allows us to compare the results of each of the touch gestures. Despite the decreasing accuracy in testing data, both classifiers still able to model selected gestures successfully. For example, it appears that gesture 'constant' and 'no-touch' is easy to recognize by both classifier, while 'pat' is mainly recognizable in RF but not in Boosting and 'stroke' is mainly recognizable in Boosting but not in RF. From all the gesture presented, 'tickle' is very difficult to classify, showing obvious confusion with scratch, for both classifier. It is partly because these two gestures performed in a similar manner, making it harder for the features having clear distinction between these two. The choice of RF as the best reported models are consistent with [10], making RF model is significantly better on combined datasets, as reported in [10]

The results for the testing set[1] are illustrated by the confusion matrices for both Random Forest and Boosting classifiers on the combined features presented in Table 6 and Table 7 respectively.

Table 5: HAART dataset: Gesture recognition rates on different feature sets

Data Set	Feature Set	Random Forest	Boosting
Training	SD	36.17%	33.50%
	BMH	22.67%	22.34%
	MSD	54.82%	53.61%
	SMMHH	65.66%	60.84%
	LBPTOP	53.01%	54.82%
	Combined	76.65%	77.67%
Testing[1]	Combined	66.53%	64.54%

Table 6: HAART testing dataset: Confusion matrix for the Random Forest classifier with an average recognition rate of 66.53%

	constant	notouch	pat	rub	scratch	stroke	tickle
constant	**34**	0	0	0	0	0	0
notouch	1	**36**	0	0	1	0	0
pat	0	0	**30**	0	1	0	2
rub	0	0	1	**16**	2	13	0
scratch	0	0	3	17	**26**	3	24
stroke	0	0	1	2	0	**18**	3
tickle	0	0	1	1	6	2	**7**

Table 7: HAART testing dataset: Confusion matrix for the Boosting classifier with an average recognition rate of 64.54%.

	constant	notouch	pat	rub	scratch	stroke	tickle
constant	**34**	1	0	0	0	0	0
notouch	1	**35**	0	0	0	0	0
pat	0	0	**19**	2	2	0	3
rub	0	0	0	**14**	2	4	0
scratch	0	0	6	12	**26**	1	20
stroke	0	0	8	7	2	**30**	9
tickle	0	0	3	1	4	1	**4**

4.2 CoST

For the CoST dataset, the data from 21 out of 31 subjects were provided for the training, and the remaining were used

[1]A mistake was found and corrected in post-hoc analysis.

Table 8: CoST dataset: Gesture recognition accuracies for different feature sets

Data Set	Feature Set	Random Forest	Boosting
Training	SD	41.24%	41.31%
	BMH	27.55%	28.88%
	MSD	44.82%	44.93%
	SMMHH	52.68%	52.56%
	LBPTOP	45.65%	46.36%
	Combined	64.52%	64.44%
Testing	Combined	59.50%	58.19%

by the challenge organizers for the testing. There were a total of 3524 and 1679 samples in the training and testing set respectively. The results for both training and testing experiments are shown in Table 8. Similar to HAART, combined features gives higher accuracy in training data compared to individual features. It yields 64.5% accuracy in RF and 64.4% in Boosting. In the testing dataset, classification of 14 gesture classes resulting achieved accuracy of 59.5% using RF and 58.2% using Boosting, which is increase by four percent than reported by Jung et al [23] (Bayes: 54%; SVM: 53%), using same database. However, direct comparison is still arguable since Jung et al using leave-one-subject-out cross-validation while our method using separate sets of testing data for evaluation purpose.

The confusion matrix in Table 9 and Table 10 respectively shows gesture such as 'massage', 'pinch', 'poke', 'press', and 'stroke' is easy to recognize by both of classifiers, having large number of correctly classified gestures than the rest. The most frequent confusion was between the following gestures: 'hit with slap', 'rub with scratch' and 'stroke; squeeze with grab'. These findings resemble those previous reported on this dataset [23].

5. CONCLUSION AND DISCUSSION

In this paper, we investigate how to automatically discriminate between different touch behaviour types that are characteristic of social affective interaction. The HAART and the CoST datasets were used for this investigation. They consist of various types of touch behaviour (e.g., patting, tapping) captured using multiple pressure sensors in a grid setting integrated into different types of substrates. Five sets of high-level features were proposed to provide a rich description of each touch gesture over time. Both individual feature and combined features were evaluated using the training dataset by cross-validations with the ensemble of learning methods Random Forest and Boosting. Within these five sets of features, the video based feature set achieved best performance on both datasets; however, the combination of all features led to better performances.

Whilst the results are well beyond chance level, various aspects could be further investigated. 1) The selection of the features and their tuning process (e.g., thresholds) could be optimized to the type of touch surface, i.e., its substrate and its cover. 2) Feature selection could be done on each set of feature in order to maximize its discriminative power; 3) Rather than using a low-level features fusion, models could be built on each set of features separately and then fused at decision level. This would allow a better exploitation of

Table 9: CoST testing datset: Confusion matrix for the Randon Forest classifier with a recognition accuracy of 59.5%.

	grab	hit	massage	pat	pinch	poke	press	rub	scratch	slap	squeeze	stroke	tap	tickle
grab	**84**	0	3	0	3	0	13	2	0	0	57	0	0	0
hit	0	**57**	0	19	0	3	0	0	0	27	0	0	22	0
massage	1	0	**88**	0	1	0	1	13	8	0	1	0	0	1
pat	0	0	0	**41**	1	1	0	0	0	0	0	1	18	5
pinch	1	2	4	2	**91**	4	9	0	1	0	8	3	1	5
poke	0	10	0	4	9	**104**	5	0	0	2	1	0	18	4
press	0	0	0	3	6	0	**84**	5	0	0	0	0	3	0
rub	3	0	5	0	0	0	0	**37**	1	0	0	13	0	0
scratch	4	0	12	2	0	0	0	23	**72**	0	0	2	0	29
slap	0	46	0	25	2	0	0	0	0	**87**	0	9	24	0
squeeze	26	0	3	1	7	0	7	4	2	0	**52**	0	0	0
stroke	1	0	3	2	0	0	1	27	10	0	0	**80**	0	2
tap	0	5	0	20	0	8	0	0	1	4	0	0	**34**	0
tickle	0	0	2	1	0	0	0	9	25	0	0	12	0	**74**

Table 10: CoST testing dataset: Confusion matrix for the Boosting classifier with a recognition accuracy of 58%.

	grab	hit	massage	pat	pinch	poke	press	rub	scratch	slap	squeeze	stroke	tap	tickle
grab	**93**	0	1	0	4	0	14	2	0	0	73	0	0	0
hit	0	**71**	0	24	1	1	0	0	0	42	0	0	29	0
massage	4	0	**94**	0	1	0	1	20	13	0	1	0	0	9
pat	0	1	0	**52**	2	2	0	0	1	0	0	1	23	6
pinch	0	1	3	3	**83**	4	4	0	1	1	6	1	2	3
poke	0	9	0	3	9	**101**	2	0	0	3	1	0	14	1
press	2	3	0	8	14	2	**92**	6	1	2	2	2	7	0
rub	3	0	5	0	0	0	0	**28**	3	0	0	9	0	2
scratch	2	0	5	1	0	0	0	17	**53**	0	0	1	0	15
slap	0	24	0	13	0	0	0	0	0	**68**	0	5	10	0
squeeze	16	0	4	0	6	0	6	3	1	0	**36**	0	0	0
stroke	0	0	5	2	0	0	1	32	15	0	0	**94**	0	5
tap	0	11	0	14	0	10	0	0	0	3	0	0	**34**	1
tickle	0	0	3	0	0	0	0	12	32	1	0	7	1	**78**

the different nature of the features and touch characteristics each captures (e.g., its shapes and its dynamics). Finally, whereas this work has focused on detecting touch types, the next step would be to model their relation with the affective or social message conveyed. This step will require integrating the modelling of types behaviour with detailed kinemtics [17, 13].

6. REFERENCES

[1] D. Atkinson, P. Orzechowski, B. Petreca, N. Bianchi-Berthouze, P. Watkins, S. Baurley, S. Padilla, and M. Chantler. Tactile perceptions of digital textiles: A design research approach. CHI '13, pages 1669–1678, 2013.

[2] N. Bianchi-Berthouze, L. Berthouze, and T. Kato. Understanding subjectivity: An interactionist view. In *User Modeling*, CISM 407, pages 3–12. Springer Vienna, 1999.

[3] N. Bianchi-Berthouze and A. Tajadura Jimenez. It's not just what we touch but also how we touch it. In *Workshop on Tactile User Experience Evaluation Methods*, CHI'14, 2014.

[4] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] J. Burgoon. Relational message interpretations of touch, conversational distance, and posture. *Journal of Nonverbal Behavior*, 15(4):233–259, 1991.

[6] J. Chang, K. MacLean, and S. Yohanan. Gesture recognition in the haptic creature. EuroHaptics'10, pages 385–391. Springer-Verlag, 2010.

[7] M. D. Cooney, S. Nishio, and H. Ishiguro. Recognizing affection for a touch-based interaction with a humanoid robot. In *IEEE International Conference on Intelligent Robots and Systems*, pages 1420–1427, 2012.

[8] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini. Tactile sensing-from humans to humanoids. *IEEE Transactions on Robotics*, 26(1):1–20, 2010.

[9] G. K. Essick, F. McGlone, C. Dancer, D. Fabricant, Y. Ragin, N. Phillips, T. Jones, and S. Guest. Quantitative assessment of pleasant touch. *Neuroscience and Biobehavioral Reviews*, 34(2):192 – 203, 2010.

[10] A. Flagg and K. MacLean. Affective touch gesture recognition for a furry zoomorphic machine. *TEI '13*, page 25, 2013.

[11] B. Friedman, P. H. Kahn, Jr., and J. Hagman. Hardware companions? what online aibo discussion forums reveal about the human-robotic relationship.

CHI '03, pages 273–280, 2003.

[12] M. Fujita and H. Kitano. Development of an Autonomous Quadruped Robot for Robot Entertainment. *Autonomous Robots*, 5:7–20, 1998.

[13] Y. Gao, N. Bianchi-Berthouze, and H. Meng. What does touch tell us about emotions in touchscreen-based gameplay? *ACM Transactions on Computer-Human Interaction*, 19(4), 2012.

[14] K. Goris, J. Saldien, B. Vanderborght, and D. Lefeber. Mechanical Design of the Huggable Robot Probo. *International Journal of Humanoid Robotics*, 08(03):481–511, 2011.

[15] K. Goris, J. Saldien, I. Vanderniepen, and D. Lefeber. The huggable robot Probo, a multi-disciplinary research platform. In *CCIS*, volume 33, pages 29–41, 2009.

[16] T. Hastie, S. Rosset, J. Zhu, and H. Zou. Multi-class AdaBoost. *Statistics and Its Interface*, 2(3):349–360, 2009.

[17] M. J. Hertenstein, R. Holmes, M. McCullough, and D. Keltner. The communication of emotion via touch. *Emotion*, 9(4):566–573, Aug 2009.

[18] M. J. Hertenstein, D. Keltner, B. App, B. A. Bulleit, and A. R. Jaskolka. Touch communicates distinct emotions. *Emotion*, 6(3):528–533, Aug 2006.

[19] G. Huisman, A. Darriba Frederiks, B. Van Dijk, D. Hevlen, and B. Krose. The TaSSt: Tactile sleeve for social touch. In *HC'13*, pages 211–216, 2013.

[20] S. Hutson, S. Lim, P. Bentley, N. Bianchi-Berthouze, and A. Bowling. Investigating the suitability of social robots for the wellbeing of the elderly. In *ACII'11*, LNCS, Vol. 6974, pages 578–587. 2011.

[21] K. Isbister, K. Höök, M. Sharp, and J. Laaksolahti. The sensual evaluation instrument: Developing an affective evaluation tool. CHI '06, pages 1163–1172, 2006.

[22] A. Jan, H. Meng, Y. F. A. Gaus, F. Zhang, and S. Turabzadeh. Automatic Depression Scale Prediction using Facial Expression Dynamics and Regression. In *AVEC'14*, pages 73–80, 2014.

[23] M. M. Jung, R. Poppe, M. Poel, and D. Heylen. Touching the void - introducing cost: Corpus of social touch. In *ICMI'14*, pages 120–127, 2014.

[24] A. Kleinsmith and N. Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, pages 1–20, 2013.

[25] M. L. Knapp and J. A. Hall. *Nonverbal Communication in Human Interaction*. Wadsworth: Cengage Learning, Bosten, USA, 7 edition, 2010.

[26] H. Meng and N. Pears. Descriptive temporal template features for visual motion recognition. *Pattern Recognition Letters*, 30(12):1049–1058, 2009.

[27] H. Meng, N. Pears, and C. Bailey. A human action recognition system for embedded computer vision application. In *CVPR workshop on Embeded Computer Vision*, pages 1–8, 2007.

[28] H. Meng, N. Pears, M. Freeman, and C. Bailey. Motion history histograms for human action recognition. In B. Kisačanin, S. Bhattacharyya, and S. Chai, editors, *Embedded computer vision*, Advances

in pattern recognition, pages 139–162. Springer, 2009.

[29] K. Nakajima, Y. Itoh, Y. Hayashi, K. Ikeda, K. Fujita, and T. Onoye. Emoballoon: A balloon-shaped interface recognizing social touch interactions. In *Virtual Reality*, pages 1–4, 2013.

[30] F. Naya, J. Yamato, and K. Shinozawa. Recognizing human touching behaviors using a haptic interface for a pet-robot. In *IEEE ICSMC*, pages 1030–1034 vol.2, 1999.

[31] Y.-W. Park, K.-M. Baek, and T.-J. Nam. The roles of touch during phone conversations: Long-distance couples' use of poke in their homes. CHI '13, pages 1679–1688, 2013.

[32] B. Petreca, N. Bianchi-Berthouze, and B. S. How do designers feel. ACII '15, 2015.

[33] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic. AVEC 2011-the first international audio/visual emotion challenge. In *ACII (2)*, LNCS, Vol. 5975, pages 415–424. Springer.

[34] B. Schuller, M. F. Valstar, R. Cowie, and M. Pantic. AVEC 2012: the continuous audio/visual emotion challenge - an introduction. In *ICMI'12*, pages 361–362, 2012.

[35] T. Shibata, K. Inoue, and R. Irie. Emotional robot for intelligent system-artificial emotional creature project. *Proceedings 5th IEEE International Workshop on Robot and Human Communication*, 1996.

[36] T. Shibata, T. Mitsui, K. Wada, A. Touda, T. Kumasaka, K. Tagami, and K. Tanie. Mental commit robot and its application to therapy of children. In *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2001.

[37] 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, 2012.

[38] D. Silvera-Tawil, D. Rye, and M. Velonaki. Interpretation of social touch on an artificial arm covered with an eit-based sensitive skin. *International Journal of Social Robotics*, 6:489–505, 2014.

[39] D. Silvera-Tawil, D. Rye, and M. Velonaki. Artificial skin and tactile sensing for socially interactive robots: A review. *Robotics and Autonomous Systems*, 63:230–243, 2015.

[40] H. Singh, M. Bauer, W. Chowanski, Y. Sui, D. Atkinson, S. Baurley, M. Fry, J. Evans, and N. Bianchi-Berthouze. The brain's response to pleasant touch: an eeg investigation of tactile caressing. *Frontiers in human neuroscience*, 8, 2014.

[41] W. Stiehl, J. Lieberman, and C. Breazeal. Design of a therapeutic robotic companion for relational, affective touch. *Robot and Human*, pages 408–415, 2005.

[42] D. S. Tawil, D. Rye, and M. Velonaki. Touch modality interpretation for an eit-based sensitive skin. In *Proceedings of IEEE ICRA*, pages 3770–3776, 2011.

[43] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic. AVEC 2014: 3d dimensional affect and depression recognition challenge. AVEC '14, pages 3–10, 2014.

[44] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic.

Avec 2013: The continuous audio/visual emotion and depression recognition challenge. AVEC '13, pages 3–10, 2013.

[45] S. Yohanan, M. Chan, J. Hopkins, H. Sun, and K. MacLean. Hapticat: Exploration of affective touch. ICMI '05, pages 222–229, 2005.

[46] 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.

[47] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. PAMI*, 29(6):915–928, June 2007.