

Touch Challenge ‘15: Recognizing Social Touch Gestures

Merel M. Jung¹, Xi Laura Cang², Mannes Poel¹, Karon E. MacLean²

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

²University of British Columbia, 2366 Main Mall, Vancouver, B.C. V6T 1Z4, Canada

m.m.jung@utwente.nl, cang@cs.ubc.ca, m.poel@utwente.nl, maclean@cs.ubc.ca

ABSTRACT

Advances in the field of touch recognition could open up applications for touch-based interaction in areas such as Human-Robot Interaction (HRI). We extended this challenge to the research community working on multimodal interaction with the goal of sparking interest in the touch modality and to promote exploration of the use of data processing techniques from other more mature modalities for touch recognition. Two data sets were made available containing labeled pressure sensor data of social touch gestures that were performed by touching a touch-sensitive surface with the hand. Each set was collected from similar sensor grids, but under conditions reflecting different application orientations: CoST: Corpus of Social Touch and HAART: The Human-Animal Affective Robot Touch gesture set. In this paper we describe the challenge protocol and summarize the results from the touch challenge hosted in conjunction with the 2015 ACM International Conference on Multimodal Interaction (ICMI). The most important outcomes of the challenges were: (1) transferring techniques from other modalities, such as image processing, speech, and human action recognition provided valuable feature sets; (2) gesture classification confusions were similar despite the various data processing methods used.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: Haptic I/O; I.5.2 [PATTERN RECOGNITION]: Design Methodology—*Classifier design and evaluation, Feature evaluation and selection*; I.5.4 [PATTERN RECOGNITION]: Applications—*Signal processing*

General Terms

Performance

Keywords

Social touch; Touch data set; Touch gesture recognition

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 2015, November 9–13, 2015, Seattle, WA, USA.

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

ACM 978-1-4503-3912-4/15/11 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/2818346.2829993>.

1. INTRODUCTION

Touch is an important non-verbal form of interpersonal social interaction; it is used to communicate emotions and social messages [8]. In order to enable artificial social agents such as robots to understand human touch input there is a need for automatic recognition of different types of touch [14]. Enabling touch-based interaction with robotic animals (e.g., [15, 18]) and humanoid robots (e.g., [6, 13]) is of interest for fields such as Human-Robot Interaction (HRI) [2]. As the recognition of touch behavior has received far less research attention than recognition of behaviors in the visual and auditory modalities (e.g., see [19]), we aimed to spark interest into this relatively new field by organizing a touch challenge.

This challenge focused on the recognition of touch gestures with social meaning that were performed by hand on a pressure-sensitive surface; we call these ‘social touch gestures’. Touch data has been collected from subjects performing different sets of touch gestures on different surfaces/embodiments (e.g., [1, 4, 5, 6, 7, 12, 13, 15]). Appropriating methods developed in more mature fields such as speech recognition and video analysis could be beneficial for moving touch recognition forward. By publicizing two distinct touch data sets, we allowed researchers with expertise in other sensory modalities to try out their processing techniques on our touch data.

The remainder of the paper is organized as follows: Section 2 describes the two provided touch data sets; Section 3 highlights the challenge protocol; an overview of the results and discussions of the test set label submissions is in Section 4; and we conclude with high-level findings in Section 5.

2. TOUCH DATA SETS

For the challenge, two data sets were made available containing labeled pressure sensor data of social touch gestures. Table 1 summarizes the data sets’ attributes.

2.1 CoST: Corpus of Social Touch

CoST [11, 12, 17] contains 14 touch gestures: grab, hit, massage, pat, pinch, poke, press, rub, scratch, slap, squeeze, stroke, tap, and tickle. These gestures were registered on an 8×8 pressure sensor grid which was wrapped around a mannequin arm. This corpus consists of the data from 31 subjects performing the 14 touch gestures in 3 variations: gentle, normal, and rough. Subjects were restricted neither in the amount of time taken for performing each gesture nor the number of gesture repetitions performed in each capture. The data provided for this challenge consisted of the

Table 1: Data set attributes as provided for the challenge.

Attribute	CoST	HAART
# of touch gestures	14	7
Sensor grid size	8×8	8×8 ^a
Sensor sample rate	135 Hz	54 Hz
Sensor values	0–1023	0–1023
Gesture duration	variable	8s ^b
Touch surface	mannequin arm	dependent on condition
Conditions	gentle and normal variations	substrates and covers
# of subjects	31	10
Train/test split	21/10 subjects	7/3 subjects
# of gesture captures	5,203	829

^atrimmed from collected 10×10 grid, ^btrimmed from collected 10s capture

gentle gesture variation (2,601 captures) and the normal gesture variation (2,602 captures). This data set was provided in both CSV and MATLAB file formats and included segmented gesture captures of varying length, sampled at 135 Hz, and containing pressure values of the 64 channels ranging from 0 to 1023 (i.e. 10 bits). Labels consisted of touch gesture type, gesture variation, and subject number.

2.2 HAART: Human-Animal Affective Robot Touch

HAART [4] contains 7 touch gestures: pat, constant contact without movement (press), rub, scratch, stroke, tickle, and ‘no touch’. These gestures were found to be the most often used of those in Yohanan et al’s Touch Dictionary [18], gathered to communicate emotion in human-animal interaction. For the HAART data set (collected from 10 subjects), each touch action was performed on a 10×10 pressure sensor [7] for 10 seconds. To assess feature robustness under realistic operating conditions when installed on a robotic animal, each subject contributed gestures with the sensor mounted on all permutations of 3 substrate conditions (firm and flat; foam and flat; foam and curve) and 4 fabric cover conditions (none; short minkee¹; long minkee; synthetic fur). The resulting data set includes 829 gesture captures (12 conditions × 7 gestures × 10 subjects minus 11 erroneous capture instances). Each capture is 10 seconds of a continuously repeated gesture, sampled at 54 Hz and trimmed to the middle 8s (432 frames); there are generally 10–15 gesture instances per capture. This data set was provided as a CSV file and included the center 8×8 frame (trimmed for consistency with CoST) with pressure values ranging from 0 to 1023. Labels consisted of touch gesture type, condition set, and subject number.

3. CHALLENGE PROTOCOL

The aim of this challenge was to develop relevant features and apply classification methods for recognizing social touch gestures. Gesture classification was independent of the condition (i.e. gesture variant, substrate and cover) for example, ‘gentle stroke’ and ‘normal stroke’ were considered to be part of the same class. Participants had the choice of working on one of the data sets or on both. For the train/test sets, subjects were randomly split into 21 train, 10 test subjects for CoST and 7 train, 3 test subjects for HAART. This

¹Minkee (or minky) is a chenille-like fabric commonly used for baby blankets and stuffed toys.

split ensured that for each of the two data sets, any one subject’s touch data belonged to either the training set or the test set.

The training data for both the CoST and HAART data set was made available to the registrants of the challenge. We provided the test data sets without class labels a month after initial publication. Participants were given 2 weeks to process the test data. Any number of test label submissions could be made up to a deadline (see Tables 2 and 3); once this date had passed, we released the true test labels as well as a summary of their results to the challenge participants.

4. TEST LABEL SUBMISSION RESULTS

Results of the test label submissions were reported in the form of a confusion matrix and accuracy was used to measure overall performance (see Tables 2 and 3).

4.1 Data pre-processing

There is not much standardization in the extraction of feature sets for touch data processing. This section discusses the data pre-processing steps that were taken by the challenge participants consisting of data filtering, feature extraction, and feature selection.

The data that was provided for the challenge was previously filtered for erroneous entries and segmented into gesture captures [12]. However, gesture captures of the CoST data set are of variable duration and can contain a single gesture instance or multiple repetitions. This increased the difficulty of automatic segmentation based on pressure differences over time. Ta et al. explored additional techniques for automatic segmentation to further reduce the amount of excess frames [16]. However, these methods for automatic segmentation did not improve classification. The gesture captures from the training set were then manually segmented based on the shape and duration which offered little to negative improvement (see Table 2) suggesting that classifiers are fairly robust to imprecise segmentation.

For the challenge, many interesting features were extracted but we describe only a couple of notable approaches here. In previous literature (e.g., [1, 7, 12, 15]) as well as for this challenge [3, 9, 10, 16] statistics were calculated from the pressure sensor data such as the mean pressure over time. Also, feature extraction methods were borrowed from other domains: speech applications, human action recognition, and image analysis. Ta et al., for example, applied the Sobel operator, an image processing technique used for edge detection [16]. By sharpening the contrast, a second set of

Table 2: Results for the CoST data set.

Paper	Classifier	Accuracy
Ta et al. [16]	random forest ^a	61.3%
Ta et al. [16]	random forest ^b	60.8%
Ta et al. [16]	SVM ^b	60.5%
Ta et al. [16]	SVM ^a	59.9%
Gaus et al. [9]	random forest	58.7%
Gaus et al. [9]	multiboost	58.2%
Hughes et al. [10]	logistic regression	47.2%
Balli Altugl et al. [3]	random forest	26.0%

^atrained on filtered data, ^btrain on all data

Table 3: Results for the HAART data set.

Paper	Classifier	Accuracy
Ta et al. [16]	random forest	70.9%
Ta et al. [16]	SVM	68.5%
Hughes et al. [10]	logistic regression	67.7%
Gaus et al. [9]	random forest	66.5%
Gaus et al. [9]	multiboost	64.5%
Balli Altugl et al. [3]	random forest	61.0%

data frames was constructed, garnering new values using the same feature extraction procedures. Most features were extracted by feature engineering however, Hughes et al. also included deep autoencoders for automatic feature extraction using dimension reduction, these features were then used to train Hidden Markov Models (HMMs) [10]. The CoST data set was used to determine HMM likelihoods for class membership; these values were included in the feature sets for both CoST and HAART data to examine the viability of applying learned features from one data set to the other.

Feature selection was performed by evaluating the performance of different features or feature sets as a whole on the training set. Relevant features selected using random forest [16] or sequential floating forward search [3] were found to improve the accuracy on the test set. Others compared the accuracies of different feature sets as a whole. Using this approach, the combination of all feature sets yielded the best results [9, 10]. Identifying a small number of highly discriminating features can benefit applications in which computational power is costly, such as on-board real-time touch recognition. Balli Altuglu and Altun [3] showed that a small feature set (number of features < 10) could perform well on the HAART data set.

4.2 Social Touch Classification

Random forest was found to be the most popular classification method [3, 9, 16] and has been used in previous work on touch gesture recognition [1, 7]. Other classification methods that were explored were Support Vector Machines (SVMs) [16], also used by [6, 11, 12], multiboost [9] (a different boosting algorithm was used by [13]), and simple logistic regression [10].

Accuracies reported for the challenge ranged from 26.0%–61.3% for the CoST data set and from 61.0%–70.9% for the HAART data set (see Tables 2 and 3). Previously reported accuracies for the CoST data set were up to 54.0% for the rough gesture variants using leave-one-subject-out cross-validation [12] and up to 64.6% when using 10-fold cross-validation [17]. For the whole CoST data set, classification independent and without knowledge of the ges-

ture variant yielded accuracies up to 52.6% using leave-one-subject-out cross-validation [12]. For the HAART data set, accuracies up to 90.3% were reported using 20-fold cross-validation when subject and condition labels were included as features [4]. However, direct comparisons between the accuracies reported for the challenge [3, 9, 10, 16] and accuracies reported by the authors of CoST [11, 12, 17] and HAART [4] are not meaningful because of the differences in data division and use of condition and/or subject information as labels.

As accuracy rates alone provide little information we looked at the confusion matrices for notable patterns. Frequent confusions between touch gestures for the CoST data set reported by the challenge participants were: ‘grab-squeeze’, ‘hit-pat-slap-tap’, ‘rub-stroke’, and ‘scratch-tickle’ [3, 9, 10, 16]. The touch gestures were difficult to distinguish across approaches for data pre-processing and classification algorithms. Previous work on the CoST data set, although using different parts and splits of data set, found similar confusions [11, 12, 17]. For the HAART data, rub and tickle were the hardest to correctly classify across challenge participant approaches [3, 9, 10, 16]. Often misclassified was rub as scratch or stroke and tickle as scratch, while the reverse (e.g., misclassification of scratch as rub) was less common. Cang et al. also found that rub and tickle were hardest to classify correctly even while using an extended version of the HAART data set [4]. Compared to the challenge results, their confusion matrices showed more symmetry, indicating there were frequent confusions among certain gesture pairs. Rub was also one of the most difficult to correctly classify for the CoST data set [3, 9, 10, 11, 12, 16, 17].

Based on observations from the recording of the HAART data set, similarities were observed in how subjects performed the touch gestures which may help to explain certain confusions. Scratch and tickle both followed a similar motion trajectory and tended to have fluttery finger movements. Rub and stroke again have analogous motions where the flat of the hand exerts pressure along a roughly linear path. Confusions between touch gestures on the CoST data set could also be explained by gestures showing similarities on characteristics such as duration, contact area, repetition probability, and frequency of direction changes [12].

5. CONCLUSION

The challenge outcomes are encouraging; participants’ various approaches open up further avenues for exploring data processing of social touch. Comparing the results from these different approaches also provided us the opportunity to pinpoint the difficulties that need to be addressed to increase the reliability of touch gesture recognition.

Feature extraction from touch data

The challenge provided insights on how techniques for feature extraction that are prominent for other modalities may be applied to touch data. Interestingly enough, many of these techniques were reasonably transferable to touch gesture data without much modification.

Future Work: While we have seen commonalities in feature sets used by challenge participants, developing a standard would help ease the feature engineering process. This challenge has allowed for the field of touch recognition to ‘pick up a few tips and tricks’ from data processing techniques used for more mature modalities, presenting an

opportunity for customizing these methods to meet the particular needs of touch data.

Difficulties in touch gesture recognition

Despite the use of different data pre-processing techniques and classification algorithms, we observed consistent classification confusions between specific gesture pairs. It is yet unclear if these classification difficulties can be resolved by finer-grained feature extraction or if the problem is actually our discretization of touch gestures. For instance, scratch and tickle could be regarded as the same gesture class.

Future Work: We suggest considering the implications of collapsing certain commonly confused gesture pairs: what defines a ‘good’ gesture set; how many gestures should comprise it; and which ones? There is a lot more to determining touch semantics and intent than performing gesture recognition, that is, the same touch gesture can be used to convey distinct social message in a different contexts. Multimodal cues could add to contextual understanding of touch data.

6. ACKNOWLEDGMENTS

This publication was supported by the Dutch national program COMMIT. The National Sciences and Engineering Research Council of Canada (NSERC) also provided partial support for this work. Finally, we are grateful to the program committee for reviewing the submitted papers and to the challenge participants whose contributions have made this challenge worthwhile.

7. REFERENCES

- [1] K. Altun and K. E. MacLean. Recognizing affect in human touch of a robot. *Pattern Recognition Letters*, 2014.
- [2] B. D. Argall and A. G. Billard. A survey of tactile human–robot interactions. *Robotics and Autonomous Systems*, 58(10):1159–1176, 2010.
- [3] T. Balli Altuglu and K. Altun. Recognizing touch gestures for social human-robot interaction. In *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, (Seattle, WA), in press.
- [4] X. L. Cang, P. Bucci, A. Strang, J. Allen, K. E. MacLean, and H. Y. S. Liu. Different strokes and different folks: Economical dynamic surface sensing and affect-related touch recognition. In *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, (Seattle, WA), in press.
- [5] J. Chang, K. E. MacLean, and S. Yohanan. Gesture recognition in the haptic creature. In *Proceedings of the International Conference EuroHaptics*, (Amsterdam, The Netherlands), pages 385–391. 2010.
- [6] 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.
- [7] A. Flagg and K. E. MacLean. Affective touch gesture recognition for a furry zoomorphic machine. In *Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction (TEI)*, (Barcelona, Spain), pages 25–32, 2013.
- [8] A. Gallace and C. Spence. The science of interpersonal touch: an overview. *Neuroscience & Biobehavioral Reviews*, 34(2):246–259, 2010.
- [9] Y. F. A. Gaus, T. Olugbade, A. Jan, R. Qin, J. Liu, F. Zhang, et al. 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), in press.
- [10] D. Hughes, N. Farrow, H. Profita, and N. Correll. 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), in press.
- [11] M. M. Jung. Towards social touch intelligence: developing a robust system for automatic touch recognition. In *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, (Istanbul, Turkey), pages 344–348, 2014.
- [12] M. M. Jung, R. Poppe, M. Poel, and D. K. J. Heylen. Touching the void – introducing CoST: Corpus of Social Touch. In *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, (Istanbul, Turkey), pages 120–127. 2014.
- [13] 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. *Robotics Research*, 31(13):1627–1641, 2012.
- [14] 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.
- [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] V.-C. Ta, W. Johal, M. Portaz, E. Castelli, and D. Vaufreydaz. The Grenoble system for the social touch challenge at ICMI 2015. In *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, (Seattle, WA), in press.
- [17] S. van Wingerden, T. J. Uebbing, M. M. Jung, and M. Poel. 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), pages 7–12, 2014.
- [18] 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.
- [19] Z. Zeng, M. Pantic, G. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.