2012 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 7-12, 2012. Vilamoura, Algarve, Portugal

Recognizing Affection for a Touch-based Interaction with a Humanoid Robot*

Martin D. Cooney, Shuichi Nishio, *Member, IEEE*, Hiroshi Ishiguro, *Member, IEEE*

***Abstract*— In order to facilitate integration into domestic and public environments, companion robots can seek to communicate in a familiar, "socially intelligent" manner, recognizing typical behaviors which people direct toward them. One important type of behavior to recognize is the displaying and seeking of affection, which is fundamentally associated with the modality of touch. This paper identifies how people communicate affection through touching a humanoid robot appearance, and reports on the development of a recognition system exploring the modalities of touch and vision. Results of evaluation indicate the proposed system can recognize people's affectionate behavior in the designated context.**

I. Introduction

In order to facilitate integration into domestic and public environments, companion robots can seek to communicate in a "socially intelligent" manner [13], recognizing typical human behaviors, such as those which convey *affection* [21]. "Affection", used in its usual sense (e.g., [12, 26, 31, 32]), refers to a positive feeling comprising love, gentleness, strong regard, and devotion; this is not the same as "affect" or emotion in general. A fundamental role in conveying affection is played by touch "behaviors", which involve not only physical touch *gestures*, but also how these gestures are perceived: i.e., their *meanings*, in terms of the degree of affection they convey.

Affectionate behaviors should be recognized because they play a vital role in establishing close relationships [1, 12], and because affection is expressed even toward non-humans such as animals and robots [31, 32]. Moreover, when affection is perceived to be lacking in the case of loneliness, consequences may include impaired sleeping, risk of suicide, and high blood pressure [3], as well as a higher incidence of Alzheimer's disease [34]. Unfortunately, many may not receive enough affection [26]. It would be useful if companion robots could provide an ancillary source of affection in people's lives.

In this work, we focus on the fundamental problem of recognizing people's affectionate touch behaviors directed toward a humanoid robot. Although many forms of robots exist, all with their own merits, a humanoid appearance was of particular interest for this study for a number of reasons: familiarity of the interface (people know how to convey various meanings to other people), people's expectation that a robot with a human-like appearance is capable of human-like communication and recognizing their behaviors [2], and previous work which has described people showing affection through touch to humanoid robots [31, 32].

Fig. 1. An example of affectionate touching (a hug) between a human (foreground right) and robot (background left)

Our requirements for an embodiment included generality and ease-of-touching: identified behavior should not be restricted to a specific robot with a unique appearance, and the embodiment should not prohibit affectionate behaviors which could be interesting or important. Humanoid robots usually have distinctive features (e.g., [7, 15, 27]), which we wished to avoid; as well, studies investigating people's behaviors toward robots usually use motionless platforms (e.g., [18, 29, 6]), partly because movements can interfere with touches [6]. Based on these considerations, a new platform was built for this study, Kirin, shown in Fig. 1 and 2, which was designed to be generic in appearance and easy-to-touch: typical for a humanoid robot in shape and size, lacking distinctive features, safe, soft, and flexible. Because motion was not desired, Kirin was not equipped with actuators, and thus may not technically be considered a robot. We refer to it hereafter as a humanoid robot "mock-up", "form" or "appearance"; nevertheless, we note that because the lack of actuators is not visible, this is only a matter of terminology: Kirin is for this study equivalent to a motionless humanoid robot.

Problematic to the current endeavor is that a countless number of touch gestures and associated meanings may exist, possibly complex: we do not know how people seek to communicate affection to a humanoid robot, or what is required for a system to recognize such behaviors. In this paper, we propose to model people's behaviors via a set of Typical Affectionate Behaviors (TAB), comprising *typical* touch gestures, their affectionate *meanings*, and their *recognizability* by a recognition system.

The rest of this paper is structured as follows. Section II discusses some related work. Section III describes our

*This research was supported by JST, CREST.

Martin Cooney, Shuichi Nishio, and Hiroshi Ishiguro are with Advanced Telecommunications Research Institute International (ATR), Hiroshi Ishiguro Lab (HIL), 2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288, Japan. Martin Cooney and Hiroshi Ishiguro are also with the Dept. of Systems Innovation, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. (E-mail: [mcooney, nishio, ishiguro]@atr.jp)

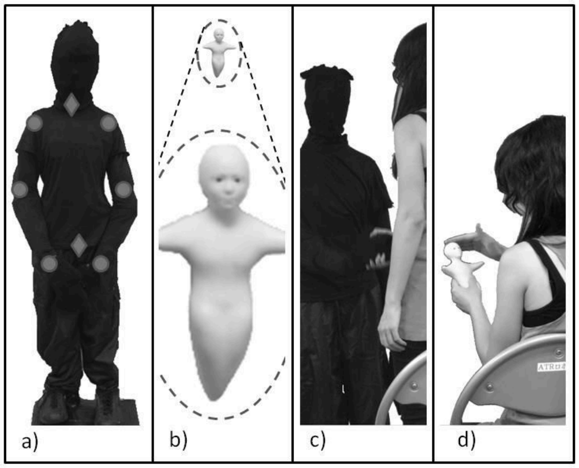

Fig. 2. Kirin and Elfoid: (a) Kirin (degrees of freedom indicated) (b) Elfoid (above: size relative to Kirin, below: magnified), (c) standing to interact with Kirin, (d) sitting to interact with Elfoid

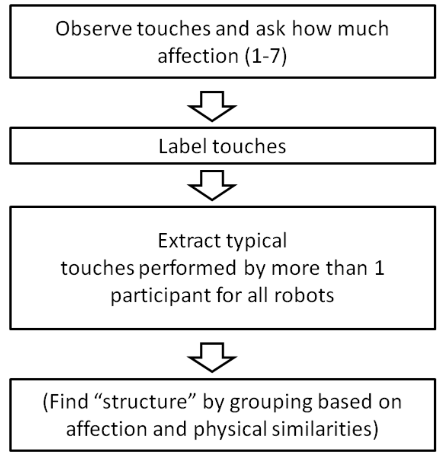

Fig. 3. Process for identifying people's typical affectionate touch behavior toward a humanoid robot

approach to modeling people's affectionate behaviors, and Section IV reports on the creation of recognition capability, which is evaluated in Section V. Section VI provides discussion, and Section VII summarizes contributions.

II. Related Work

Accounts of people touching robots with affection can be found both in the literature (e.g., "Trisha" hugging Kismet in [32]) and in popular culture (e.g., John Connor bonding with a humanoid robot by teaching it how to "high-five" in "Terminator II".) Benefits have been found from engaging in such touch-based and affectionate interactions with robots. A "Hug Machine" was constructed in 1965 to administer calming touch [10], and pioneering work with companion robots Paro and AIBO showed that touching robots can reduce stress, improve relationships, and otherwise contribute to health [17, 33].

How people touch a robot has been partly addressed. Basic types of touches such as "rub" or "pat" have been proposed [23], as well as more complex 3D touches such as "pinch" and "twist" performed toward a humanoid robot [36]. Within the context of play, one study observed and manually labeled touches to find "interaction categories" for playing with a child-sized humanoid robot [24]; also, Knight and her colleagues provided a useful list of play gestures people performed toward a teddy bear robot [18]. An approach for automatically clustering generic touch data was described in [30], but interpreting such results remains problematic. Thus, previous studies have not investigated how people touch a humanoid robot outside the context of play.

Another challenge involves knowing how people perceive the *meanings* of their touches; without such knowledge, a designer may be required to make arbitrary or suboptimal decisions for how a robot should react to each touch. Thus, Francois et al. analyzed the summed output from five force sensors on a dog-like robot to distinguish between gentle and strong touch [8]. Similarly, an interesting work by Stiehl et al. sought to distinguish pleasant or painful touches toward a humanoid robot's arm segment [29]. For the current problem, the effect of location on touch should also be considered, and the categorization of meaning should not be limited to a binary case, as many kinds of touches exist.

Toward recognizing touches, the question of which modality should be used has also not received attention. Although touch sensors have been used in many works (such as those above, and others discussed in [19]), touch sensors can be difficult to manufacture and maintain, portability between robots may not be straight-forward, and useful information may be lost (e.g. postural information during the preparation or retraction phase of a gesture). One interesting alternative is suggested by recent work in psychology [14], which found that people can interpret the meaning of touch behavior simply by observing it. Thus, computer vision, with its own rich literature, could be used to recognize touch behavior. Vision sensors may also however experience problems, e.g., with occlusions, varying illumination, and detecting when a person is touching the robot. Therefore, both sensor modalities (touch and vision) should be compared.

In summary, the main novel contributions of this work are three-fold: revealing how people typically touch a humanoid robot appearance, finding what touch gestures mean in terms of a non-binary degree of affection, and exploring the suitability of the modalities of touch and vision for recognizing touches (including finding useful features). This knowledge, summarized as a set of typical affectionate behaviors, TAB, is intended to allow the engagement of a companion robot in affectionate interactions, toward facilitating bonding and successful integration into everyday human environments.

III. MODELING PEOPLE'S AFFECTIONATE BEHAVIOR

Data collection was conducted in two phases: first to reveal how people touch a humanoid robot form, and second to build a recognition system (described in Section IV).

A. Goal

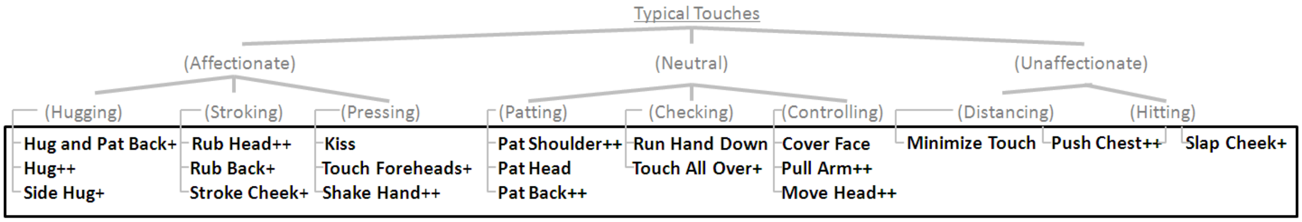

Fig. 4. Typical touches toward a humanoid robot appearance organized by affection and type of touch; "+" symbols to the right of a typical gesture indicate typicality: "++" very frequent, "+" frequent

The main goal of this section was to find *typical* touches toward a humanoid robot. Typical touches were sought in order to avoid outliers arising from idiosyncrasies in individual participants' behavior patterns and the appearances of specific robots. Thus, "typical" was defined here as "performed by more than one person for each robot form". The affectionate meaning of gestures was also sought, in order to understand how touches interrelate.

B. Participants

Data were obtained from 21 participants (9 females and 12 males; 19 Japanese and 2 non-Japanese; average age=24.1 years, SD=4.4 years).

C. Procedure

The literature did not indicate a way to find people's typical affectionate behaviors, TAB. The approach chosen is shown in Fig. 3. Participants were asked to touch robot forms in various ways and to themselves indicate how much affection they would feel if touched in this way. From this, typical touch patterns could be identified.

Thus, participants were first admitted by the experimenter to an empty room with two robot mock-ups (Kirin and Elfoid, shown in Fig. 2), a chair, and a desk; the door was closed for privacy. Once seated, participants were asked to read a handout containing instructions ("You will be asked by the experimenter to convey various intentions and emotions through touch"), as well as a list of intentions which a person might typically wish to convey through touch to a humanoid robot (e.g., "greeting"). "Affection" was described as in Section I. Then, participants touched Kirin and Elfoid in counterbalanced order, while standing or sitting respectively. During the trials, the experimenter noted in writing how participants touched the robots and asked participants to say what they had done in their own words when there was ambiguity; video footage was also acquired. After each touch, participants described the degree of affection conveyed by the touch on a 7 point scale (1 no affection at all, 4 regular, 7 much affection). Finally, short interviews were conducted. During sessions, the experimenter was the only other person present. Sessions lasted ~30 minutes.

D. Conditions

Robot

Kirin: Kirin (shown in Fig. 2a and c) was designed for this study to be typical for a humanoid robot in shape and size, neutral (lacking in distinctive features), and safe, soft and flexible (easy-to-touch). Humanoid robots mostly have arms and legs, and many are the size of an adult human (e.g. Robonaut [7], Geminoid [27], and ASIMO [15]), hence Kirin was made to appear mostly human-shaped, with a height of 168cm (between average male and female height). Distinct features and hard portions of its mechanical frame (articulated at the shoulders, neck, waist, elbows, and wrists) were covered with dark material, also making Kirin soft and easy to touch. Kirin lacks actuators and sharp edges, and has a stable base, to allow safety and robustness even when touched firmly.

Elfoid: Unclear was the extent to which people's behavior toward Kirin would coincide with behavior toward robots with a different appearance; therefore, an additional mock-up was selected, Elfoid [22], (shown in Fig. 2b and d), which was different in appearance from Kirin: abstractly-shaped (no hands or legs), small, and creature-like with eyes and a nose.

Reasons for Touching

Requests from initial users in free-interaction pre-trials also indicated that reasons were required for touching a robot, without which users had difficulty imagining what to do. For this, a list of typical reasons for touching were assembled based on sources in human science [14, 16] (described in Appendix A); e.g., showing "love", "anger", "support", "thanks", "fear", or a "greeting". As a result, a variety of different touches could be observed.

E. Processing

After the trials, video footage was checked for completeness, notes were codified with short gesture labels (920 in total, 239 unique) by the experimenter, and these labels were input to a database for processing along with affection scores normalized to fall within the interval of [0, 1]. Outliers were removed, reducing the number of gestures to 44 and 45 for Kirin and Elfoid; then gestures specific to only one robot form were removed, leaving 20 typical gestures. Among the filtered gestures were six gestures with significant affectionate meaning: "High Five" (.83) and "Shove Shoulder" (.08) were typical only for Kirin, whereas "Hug and Rub Head" (.96), "Rub Belly" (.83), "Turn Upside-down" (.06) and "Throw Robot" (.04) were typical only for Elfoid. "High Five" and "Shove Shoulder" were not performed for Elfoid because its arms are not articulated and it does not have shoulders; "Turn Upside-down" and "Throw Robot" could not be performed on Kirin due to its large size. Last, touches were grouped by affection value and manner of touching.

F. Results

Twenty members of the set of typical affectionate behavior, TAB, were identified which seem useful to recognize in order to gauge a person's affection (these are shown in the rectangle at the bottom of Fig. 4). The top part of the figure shows that the most affectionate types of gestures were hugging, stroking, and pressing gestures such as kissing. Neutral gestures involved patting, checking, and controlling, whereas unaffectionate gestures included hitting and distancing.

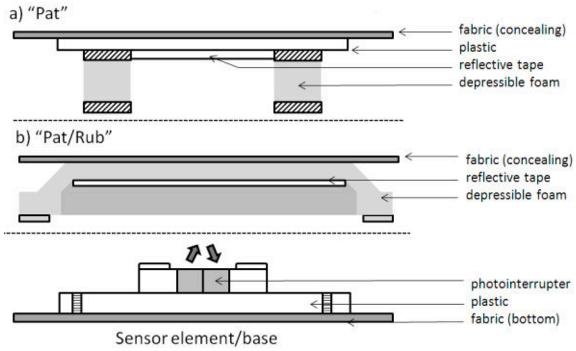

Fig. 5. Simple touch sensors built

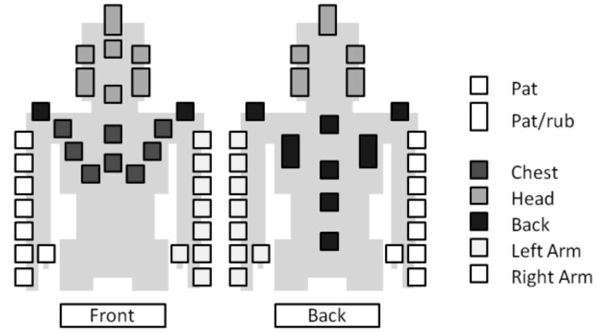

Fig. 6. Touch sensor placements

Touches exhibited complex traits (indicated below via italics). "Hug and Pat Back" was a *simultaneous* touch in which a person hugged the robot form while patting its back. "Side Hug" was a *side-initiated* gesture involving first moving adjacent to and then reaching an arm over a robot form's shoulders. Stroking gestures such as "Rub Head" or "Rub Back" were typically *along the fur* (up to down, or front to back). "Minimize Touch" (e.g., a quickly retracted poke) and "Cover Face" were *soft but not affectionate*, confirming that softness alone cannot be used to recognize affection ("Push Chest" was also sometimes soft (moving the robot away) and sometimes hard (a shove)). "Shake Hand" exhibited *high variability* and was performed in many ways, involving one or two hands from the participant and robot form. In addition, *preparatory* touches were often observed (e.g., touching Kirin's shoulder briefly before performing a gesture, or holding Elfoid). (Holding touch behavior toward a robot was reported previously in [18]).

The interviews gave insight into how participants perceived the two humanoid robot forms. Kirin was described by two participants as being like a human or a friend, whereas five participants claimed Elfoid resembled a baby or child, and two likened Elfoid to a doll, or a small animal or pet. Eleven participants said Kirin was easiest to use and that they could express themselves better with it; six said this was because of its size, three said because of its human-likeness, and one said it was because of its shape and articulation. Five participants said Elfoid was harder to touch because it was smaller, and three said it was because it was not articulated. One user however said he was "freer" when using the Elfoid (because he could lift it). Thus, some people did perceive the two robot forms as being human-like, which may explain why the identified touches are mostly familiar within the context of human-human touching.

Comparison was made with other work. From the list of play gestures found by Knight and her colleagues [18], "handshake", "head pat", and "shoulder-tap" were observed, but tickling, foot-rubbing, or care-taking were not. For Noda et al.'s list of interaction categories [24], analogies were found for "Let's shake hands", "Give me a hug", and "I wish you'd pat me on the head" but not for more abstract categories such as "Do you think I'm cute?" or "Whee!" These shared observations support that, although some differences emerge for specific contexts, typical behavior exists, and should be recognized.

IV. RECOGNIZING AFFECTION

How can people's affectionate touches be recognized by a humanoid robot? Both touch and vision-based approaches, or a combined approach, could be possible; therefore, different versions of a recognition system were implemented using Kirin as a platform, due to participants' comments from the previous section that it was easiest to use.

A. Modality for recognition

TOUCH

We posited that touches can be separated into two basic types: perpendicular ("pats") and lateral ("rubs"). Therefore two sensor types were built, one to register pats and the other to detect pats or rubs. The concept for both sensors comes from [25]. Each sensor consists of a housing and a photo-interrupter element inside. When a person touches the sensor housing, it depresses, causing more light from an infrared LED to be reflected back to a photo-transistor and resulting in a measurable variance in voltage.

Unclear was how to design the sensor housings and attach them to a robot. Two sensor housings were designed, shown in Fig. 5, over SHARP GP2S60 photointerrupters, which were attached to Kirin in the form of a sensor suit, for portability. Designs were informed by requirements for touch sensors stipulated in previous studies: sufficient area coverage and sensitivity, ease of manufacture, softness and lightness, toughness, and conformability to curved surfaces [4, 25]. The resulting sensors were large (approx. 6 X 6 and 15 X 6cm) in order to capture most touches, soft (plastic and soft low-resistance foam) and light (approx. 11g and 21g), but robust enough to be shoved or slapped without breaking. The sampling rate to read all sensors was approximately 20Hz (50ms), sufficient for online recognition, and the sensors could register sufficiently light and strong touches for the designated task. Placement locations (shown in Fig. 6) were informed by the data acquisition of the previous section. Eight sensors each were placed at five areas—chest, face, back, and both arms—resulting in a total of forty touch sensors, which is a typical number (Omron's NeCoRo, AIBO, and Paro have few sensors [9, 33], and Sensate Bear and Haptic Creature have fifty to sixty sensors [18, 35]).

VISION

One popular approach toward gesture recognition (e.g., [20]) involves obtaining postural information (skeletons) from

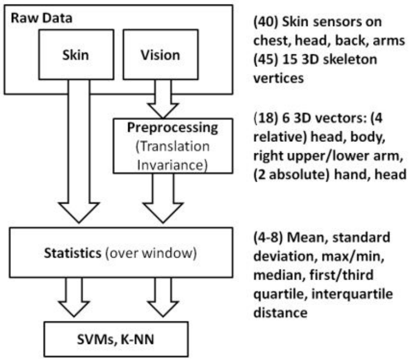

Fig. 7. Feature calculations for SVMs and k-NN

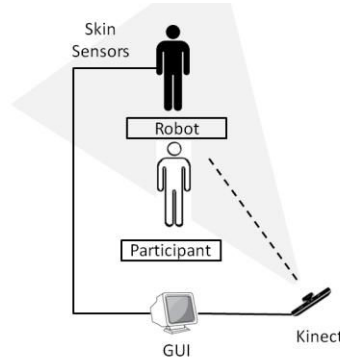

Fig. 8. Setting for data acquisition: skin sensor and Kinect data are recorded while a participant touches Kirin

people, e.g., using motion capture, but setting up multiple cameras in a sufficiently wide space may not be practical, and markers can interfere with natural touching. To avoid such problems, Microsoft Kinect was selected. A development kit requiring no calibration was first used, but problems of false positives (jumping of skeleton vertices) and the robot frequently being recognized as a person led us to install Open NI/Prime Sense, which required some calibration. The sensor was positioned roughly 2.3m away (just enough to capture the length of a tall person's body from head to feet), and behind and to the right to facilitate model fitting while ensuring right (dominant) hand visibility.

B. Recognition Technique and Features

Recognizing typical affectionate behavior from the sensor data was broken down into two sub-problems: classifying gestures and estimating affection values. For the former, in our previous work [6], Support Vector Machines (SVMs) were used. The k-Nearest Neighbor algorithm (k-NN) was also implemented for comparison, as it gives consistently good results, at times surpassing more complicated techniques in performance (e.g., [37]). For the latter case, Support Vector Regression (SVR) was selected, as this technique has been used for a similar problem of estimating the emotional content of sounds (e.g., [11]).

Set-up was as follows. For SVMs and SVR, "LIBSVM -- A Library for Support Vector Machines" [5] was used, obtaining parameters C=32.0, γ=.03125 for touch, C=8.0, γ=.125 for vision and C=8.0, γ=.03125 for touch/vision. For k-NN, k=3 was selected. Feature selection was guided by modality-specific requirements, as shown in Fig. 7. In general, touch-based recognition involves consideration of location, force, and temporal/frequency information (e.g., [18, 8, 23, 30]); such information, encapsulated in values for each skin sensor, was fed to the SVMs and k-NN after lessening the feature volume by calculating temporal statistics features found to perform well in [6] over a short window of several seconds.

Preprocessing was conducted on selected raw skeleton vertex features obtained from the Kinect to achieve translation invariance, as is common in vision-based recognition (e.g., [20, 28]). For the current problem, the absolute height of a person's hand and head were also expected to hold some meaning (e.g., regarding the part of a robot being touched and the part of the human touching). These features were also further processed to yield 8 statistics per feature for input to SVMs and k-NN. For the combined touch/vision system, fusion was attained by simply combining both forms of features ("early fusion"). Thus, 320, 144, or 464 features were delivered per gesture for the touch, vision, and touch/vision systems.

C. Affectionate Touch Data Acquisition

Data were acquired from 17 participants (10 females and 7 males; 8 Japanese and 9 non-Japanese; average age 31.8 years, SD=6.3 years), who were asked to perform the twenty typical affectionate behaviors.

Fig. 8 shows the setting for data acquisition. After receiving instructions regarding purpose, sensors used, and the list of gestures to perform, participants received a mask to cover their lower faces (for privacy and hygiene when kissing the robot) and stood in front of Kirin.

Before recording, participants were allowed to practice touching Kirin. For convenience, participants were asked to touch with their right hand or both hands (for left-handed users, the Kinect could be shifted to the left side with little loss of generality). To record gestures, participants performed a calibration pose for the Kinect, touched the robot (prompted by a WAV file), returned to the calibration pose when they felt they were done, and described how much affection they would feel if someone touched them in this way using a 7-point scale. Participants were not told where touch sensors were placed (although locations could be identified through touch or visual examination). A GUI was used by the experimenter to facilitate tasks such as randomization of gesture orders, recording, WAV file playback, label insertion, real-time viewing and playback of sensor data (skin sensor, skeleton and depth data), and counting recordings. 20 gestures were recorded 5 times per participant, resulting in approximately 1700 skin, skeleton, and depth files.

V. EVALUATION

The different versions of the system (using touch, vision, or both; and SVMs or k-NN) were compared with regard to recognition capability via leave-one-out cross-validation. Recognition capability was evaluated in terms of accuracy (the number of gestures recognized correctly divided by the total number). Likewise, linear error for estimating affection was calculated for each fold, and then averaged.

Results for classification accuracy, shown in Table I, were surprising in three ways. First, the top-performing system performed well (91%) despite the difficulty of the problem

	Kiss	Hug and Pat Back	Hug	Touch Foreheads	Stroke Cheek	Side Hug	Rub Back	Rub Head	Shake Hand	Pat Back	Run Hand Down	Pat Shoulder	Pat Head	Pull Arm	Move Head	Touch All Over	Cover Face	Minimize Touch	Push Chest	Slap Cheek
Kiss	92	0	0	3	0	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0
Hug and Pat Back	2	81	16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Hug	0	17	82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Touch Foreheads	1	1	3	93	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
Stroke Cheek	0	0	0	0	93	0	0	0	0	0	0	0	0	0	3	1	1	1	0	1
Side Hug	0	0	0	0	0	98	0	0	0	1	0	1	0	0	0	0	0	0	0	0
Rub Back	1	0	1	2	0	0	92	0	0	4	0	0	0	0	0	0	0	0	0	0
Rub Head	0	0	0	0	0	0	0	90	0	0	0	0	5	0	1	1	3	0	0	0
Shake Hand	0	0	0	0	0	0	0	0	96	0	0	0	0	4	0	0	0	0	0	0
Pat Back	0	1	0	0	0	0	3	0	0	94	0	0	0	0	0	1	0	1	0	0
Run Hand Down	0	0	0	1	0	0	0	0	0	0	87	1	1	2	4	4	0	0	0	0
Pat Shoulder	0	0	0	0	0	0	0	0	0	0	3	94	0	0	0	1	0	1	0	1
Pat Head	0	0	0	0	0	0	0	4	0	0	1	1	94	0	0	0	0	0	0	0
Pull Arm	1	0	0	1	0	0	0	0	4	0	0	0	0	91	0	2	0	1	0	0
Move Head	0	1	0	0	7	1	0	3	0	0	0	1	0	0	83	0	3	1	0	0
Touch All Over	0	1	1	1	0	1	3	0	1	0	3	1	0	2	5	79	1	0	1	0
Cover Face	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	1	92	1	1	0
Minimize Touch	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	93	4	0
Push Chest	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	95	0
Slap Cheek	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	1	0	1	1	94

Fig. 9. Confusion Matrix for SVMs and Touch/Vision: rows represent ground truth, columns are classifier output; darkly shaded cells indicate good performance, lightly shaded cells indicate poor performance, and bordered cells indicate confusion.

(considerable variance in people's gestures and similarities between some gestures) and the simplicity of our approach (using either hand-crafted sensors, or a single off-the-shelf sensor). Second, the vision-based system performed better than expected in relation to the touch-based system (+5.6%), despite requiring much less development time. Third, knowledge from each modality was complementary, with a clear distance between the vision and fused system (+13.0%). k-NN results, although lower, confirmed this trend.

The touch-based system performed well for some gestures such as "Push Chest" (88%) and "Pat Head" (85%), but had trouble with recognizing "Kiss" (45%), "Touch All Over" (61%), and "Hug and Pat Back" (64%). The main difficulty resulted from light gestures performed at different locations; e.g., participants kissed the robot form's forehead, cheek, or mouth. Likewise, "Minimize Touch" led to a high rate of false negatives for "Shake Hand" (17%), "Touch Foreheads" (12%), "Kiss" (11%), and "Pat Shoulder" (11%). "Stroke Cheek" was confused with "Slap Cheek" (13%), probably because some participants slapped lightly. Compared to the vision-based system, touch did better at identifying "Push Chest", "Hug", and "Rub Back", as shown in Table II.

The vision-based system performed well for "Slap Cheek" (92%), "Side Hug" (92%), and "Shake Hand" (88%), but less well for "Hug and Pat Back" (63%), "Move Head" (65%), and "Touch All Over" (72%). "Stroke Cheek" was confused with "Cover Face" (14%) and "Move Head" (9%), because these gestures involve a prolonged touch to the robot form's face; in such cases the Kinect sensor had trouble detecting where the participant's hand ended and the robot form's head began, leading to imprecision. For "Hug and Pat Back", confusion occurred with "Hug" (30%), which is natural because the sensor could not see the participant's hand. Compared to the touch system, touches distinguishable by touching body part (e.g., "Kiss" or "Touch Foreheads") or preparatory motion ("Side Hug" or "Slap Cheek") could be better recognized.

The combined touch-vision system did well for "Side Hug" (98%), "Shake Hand" (96%), and "Push Chest" (95%), with lowest results for "Touch All Over" (79%), "Hug and Pat Back" (81%), "Hug" (82%), and "Move Head" (83%).

TABLE I. RECOGNITION ACCURACY BY MODALITY AND RECOGNITION TECHNIQUE

	Touch	Vision	Both
SVMs	71.9%	77.5%	**90.5%**
k-NN	63.3%	67.4%	82.4%

TABLE II. COMPARING TOUCH AND VISION MODALITIES (SVMS): LARGEST IMPROVEMENTS IN RECOGNITION ACCURACY OBSERVED

Vision	Touch
Kiss (+41%)	Push Chest (+12%)
Side Hug (+26%)	Hug (+9%)
Touch Foreheads (+16%)	Rub Back (+8%)
Shake Hand (+15%)	Cover Face (+8%)
Slap Cheek (+14%)	Pat Head (+6%)

Confusion existed between "Hug and Pat Back" and "Hug" (17%), "Move Head" and "Stroke Cheek" (7%), and "Rub Head" and "Pat Head" (5%). This, and similar affection scores, indicate that it might be desirable to merge "Hug and Pat Back" and "Hug" into a single gesture if a context does not require the distinction. The confusion matrix in Fig. 9 offers additional details.

Last, average error for estimating affection using SVRs was also calculated and found to be 0.13. This shows that a robot can estimate affection in a touch interaction even without making assumptions regarding gestures.

In summary, the results indicated that if development time is constrained, a robot's location is known, and occlusions are not a problem, vision can be an option to consider. However, if resources are available, a more "human-like" solution of using both modalities can yield best recognition performance, by combining the strengths of either modality.

VI. DISCUSSION

A. Completed System

Data analysis revealed twenty typical touches which can be used to model people's affectionate behavior. A system was then built based on the modalities of touch and vision to recognize the touches with an accuracy of 91%. These results are summarized in Table III. The reader is also referred to the accompanying video. (N.B., affection values differed slightly in some cases from initial estimates found in Section III which were based on a variable number of participants.) With the acquired knowledge, a companion robot can recognize people's behavior and make decisions based on perceived affectionate meaning. E.g., a robot can recognize if it is hugged and reciprocate, or choose a different response with a suitable degree of affection. Through vision, a robot can interpret behavior directed toward a different robot or even seek to interpret human-human behavior (non-applied physical gestures could also be detected). In the presence of occlusions, a robot can rely on its sense of touch. This will allow for richer, more enjoyable affectionate interactions which we hope will facilitate the introduction of companion robots into daily environments.

TABLE III. TAB (TYPICAL AFFECTIONATE BEHAVIORS)

Gesture	Affection[1]	Accuracy[2]	Gesture	Affection[1]	Accuracy[2]
Kiss	0.95 (0.076)	92%	Run Hand Down	0.60 (0.22)	87%
Hug and Pat Back	0.93 (0.087)	81%	Pat Shoulder	0.57 (0.16)	94%
Hug	0.92 (0.091)	82%	Pat Head	0.56 (0.14)	94%
Touch Foreheads	0.82 (0.16)	93%	Pull Arm	0.50 (0.19)	91%
Stroke Cheek	0.81 (0.11)	93%	Move Head	0.38 (0.21)	83%
Side Hug	0.76 (0.14)	98%	Touch All Over	0.29 (0.16)	79%
Rub Back	0.75 (0.14)	92%	Cover Face	0.24 (0.21)	92%
Rub Head	0.75 (0.18)	90%	Minimize Touch	0.085 (0.11)	93%
Shake Hand	0.66 (0.10)	96%	Push Chest	0.078 (0.10)	95%
Pat Back	0.60 (0.085)	94%	Slap Cheek	0.039 (0.16)	94%

1 Average affection values (and standard deviations) are based on the responses of the 17 participants in Section IV.C

2 Recognition accuracies are for the SVM Touch/Vision system.

B. Generality

Limitations of the current exploratory study should be recognized, with respect to platform, approach, and target group (age and culture). We cannot know everything that people will do toward a robot based on this study alone. E.g., specific motions by a humanoid robot could elicit certain touch behaviors (e.g., a robot raising its arm could result in an increased observation of “High Five”). As well, observances of typical touches were affected by the list of reasons provided (e.g., “Move Head” might not be observed if a person is not expected to touch a robot to cause it to move).

Nonetheless, we feel the current results do give some idea of what to expect. Initial experimentation with a third humanoid robot form yielded few unique gestures not shown to Elfoid or Kirin; as well, we expect speech capability in more complex robots will not stop people from touching them, because people touch other people. Although the approach used for asking people to say how affectionate their touches had been is novel, the averages and standard deviations in Table III show some consistency in how touches were perceived. Regarding target group, Hertenstein et al. found that people from cultures with different expected levels of intimacy (America and Spain) performed touches in a very similar way [14]. Yet, more than anything, the behaviors identified are mostly also performed by humans toward other humans; intuitively we can expect to see these fundamental touches performed as well toward a human-like robot. Therefore, we expect that the identified gestures can provide a reasonable estimation of what might be seen in other contexts.

C. Future Work

Next steps will include demonstrating the usefulness of recognizing affection in a practical application. The effect of a robot’s kinesics (posture and motions) on people’s behavior, finer “granularity” of recognition (e.g., recognizing relatively affectionate or unaffectionate handshakes) and coarser recognition of gesture categories will extend the current work. As well, strategies for safely and successfully performing affectionate behaviors such as hugs or side hugs toward a moving human, possibilities for affectionate memory/learning, and the manner in which affectionate behavior is expressed in other modalities (e.g. speech or vision) and toward other targets (e.g., non-humanoid forms) as well as for specific robots and contexts, will be explored.

VII. CONCLUSION

Observing people’s touches toward a humanoid robot form revealed a confusing and complicated world in itself. Yet, we were able to find twenty typical touch patterns which carried significant affectionate content, ranging from hugging, stroking, holding, and kissing, to hitting or distancing behaviors. A recognition system based on Support Vector Machines (SVMs) and Support Vector Regression (SVR), simple statistical features, and touch/visual data was developed to recognize affectionate behavior. Recognition accuracy was found to be highest when using both the modalities of touch and vision: 90.5%. As well, affection could be estimated with an average error of 0.13. The observed results suggest how a humanoid robot system can recognize people’s affection in a touch-based interaction.

APPENDIX

Reasons for touching a humanoid robot were required to observe a variety of touches from participants, but were not known from previous studies. We considered that humanoid robots are a) robots which are b) similar to humans but c) tend to be more limited in terms of communicative capabilities. Therefore, we merged three sources: a) some reasons why people might touch a robot or machine (inspecting or getting it to move), b) a list of typical reasons why people touch other people [16], and c) a list of fundamental emotions conveyable through touch, which people will likely try to express toward robots with restricted communication capabilities [14]. This resulted in the list of typical reasons shown in Table IV, through which various touches could be observed.

ACKNOWLEDGMENT

We'd like to thank Takashi Minato for help with the skin

sensors, and everyone else who supported this project.

REFERENCES

[1] J. Bowlby, "The nature of the child's tie to his mother," *International Journal of Psychoanalysis* 39 (5): pp. 350–73. 1958.

[2] A. G. Brooks and R. C. Arkin, "Behavioral overlays for non-verbal communication expression on a humanoid robot," Auton. Robots 22(1): 55-74, 2007.

[3] J. Cacioppo and W. Patrick, "Loneliness: Human Nature and the Need for Social Connection," New York: W.W. Norton & Co., 2008.

[4] G. Cannata, M. Maggiali, G. Metta, and G. Sandini, "An Embedded Artificial Skin for Humanoid Robots," *MFI 2008*, pp. 434-438, 2008.

[5] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM TIST*, 2:27:1-27:27, 2011.

[6] M. D. Cooney, C. Becker-Asano, T. Kanda, A. Alissandrakis, and H. Ishiguro, "Full-body Gesture Recognition Using Inertial Sensors for Playful Interaction with Small Humanoid Robot," *IROS 2010*, 2010.

[7] M. A. Diftler, R. Platt, Jr., C. J. Culbert, R.O. Ambrose, and W. J. Buethmann, "Evolution of the NASA/DARPA Robonaut Control System," in *ICRA 2003*, 2003.

[8] D. Francois, D. Polani, and K. Dautenhahn, "On-line behaviour classification and adaptation to human-robot interaction styles," HRI07, pp. 295 – 302, 2007.

[9] M. Fujita, "On activating human communications with pet-type robot AIBO," *Proceedings of the IEEE*, Vol. 92 (11), pp. 1804-1813, 2004.

[10] T. Grandin, "Calming Effects of Deep Touch Pressure in Patients with Autistic Disorder, College Students, and Animals," *Journal of Child and Adolescent Psychopharmacology* 2 (1): pp. 63–72, 1992.

[11] B. Han, R. Dannenberg, and E. Hwang, "Smers: Music Emotion Recognition Using Support Vector Regression," *10th. International Society for Music Information*, 2009.

[12] H. Harlow, "The Nature of Love," *American Psychologist* 13 (12): pp. 573–685, 1958.

[13] M. Heerink, B. Kröse, B. Wielinga, and V. Evers, "Enjoyment, Intention to Use And Actual Use of a Conversational Robot by Elderly People," HRI 2008, 2008.

[14] M. J. Hertenstein, D. Keltner, B. App, B. Bulleit, and A. Jaskolka, "Touch communicates distinct emotions," *Emotion*, 6, 528-533, 2006.

[15] K. Hirai et al. "The Development of Honda Humanoid Robot," *ICRA* 1998.

[16] S. E. Jones and A. E. Yarbrough, "A naturalistic study of the meanings of touch," *Communication Monographs*, 52, 19-56, 1985.

[17] M. Kanamori et al., "Pilot study on improvement of quality of life among elderly using a pet–type robot," *IEEE CIRA*, pp. 107–112, 2003.

[18] H. Knight, R. L. Toscano, W. D. Stiehl, A. Chang, Y. Wang and C. Breazeal, "Real-time social touch gesture recognition for sensate robots," *IROS 2009*: 3715-3720, 2009.

[19] M. H. Lee and H. R. Nicholls, "Tactile sensing for mechatronics-a state of the art survey," *Mechatronics*, Elsevier, Vol. 9, pp. 1-31, 1999.

[20] C. Li, S. Q. Zheng and B. Prabhakaran, "Segmentation and recognition of motion streams by similarity search," *TOMCCAP* Vol. 3 (3), 2007.

[21] K. McGee, "A touch of the future: contact-expressive devices," Multimedia, IEEE, 11 (1), 8-11, 2004.

[22] T. Minato, S. Nishio, K. Ogawa and H. Ishiguro, "Development and Study on a Cellular Phone-Type Teleoperated Android, 'Elfoid'," *RSJ*, 302-4, 2011 (in Japanese).

[23] F. Naya, J. Yamato and K. Shinozawa, "Recognizing Human Touching Behaviors using a Haptic Interface for a Pet-robot," *Proc. SMC'99*, pp. II-1030-1034, 1999.

[24] T. Noda, T. Miyashita, H. Ishiguro and N. Hagita, "Map acquisition and classification of haptic interaction using cross correlation between distributed tactile sensors on the whole body surface," *IROS 2007:* pp. 1099-1105, 2007.

[25] Y. Ohmura, Y. Kuniyoshi and A. Nagakubo, "Conformable and Scalable Tactile Sensor Skin for Curved Surfaces," *ICRA*, pp. 1348-1353, 2006.

[26] J. W. Prescott, "Body pleasure and the origins of violence," in "The Bulletin of The Atomic Scientists," pp. 10-20, November 1975.

[27] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro and N. Hagita, "Android as a Telecommunication medium with Human-like Presence," *HRI 2007*, 2007.

TABLE IV. REASONS FOR TOUCHING A HUMANOID ROBOT

Meaning for Touch	Basic Concept	Source
(Support/Sympathy)	"Poor robot, it's okay"	[14, 16]
(Appreciation)	"Thanks!"	[16]
(Inclusion)	"We're friends"	[16]
(Physical Attraction)	"How handsome/cute!"	[16]
(Love)	"I love you"	[14, 16]
(Seek Support/Sympathy)	"I'm tired, my head/belly/throat hurts"	[16] *
(Seek Love)	"I'm lonely/sad"	[16] *
(Playfulness)	"Let's play! Hey!"	[16]
(Control)	"Move body/arm/head"	[16] *
(Greeting)	"Hello/Goodbye"	[16]
(Task-related)	"Inspect the robot"	[16] *
(Emotion: Angry)	"I hate you"	[14]
(Emotion: Disgust)	"Ewww, gross"	[14]
(Emotion: Fear)	"I'm scared"	[14]

* Interpreted for the context of touches directed toward a humanoid

[28] M. Sigalas, H. Baltzakis and P. E. Trahanias, "Gesture recognition based on arm tracking for human-robot interaction," *IROS 2010*: pp. 5424-5429

[29] W. D. Stiehl and C. Breazeal, "Affective Touch for Robotic Companions," *ACII 2005*: pp. 747-754: 2004.

[30] T. Tajika, T. Miyashita, H. Ishiguro and N. Hagita, "Automatic Categorization of Haptic Interactions-What are the Typical Haptic Interactions Between a Human and a Robot?" *Humanoids'06*, 2006.

[31] S. Turkle, "A Nascent Robotics Culture: New Complicities for Companionship," AAAI Technical Report Series, 2006.

[32] S. Turkle, C. Breazeal, B. Scasselati and O. Dasté, "Encounters with Kismet and Cog: Children's Relationships with Humanoid Robots," Humanoids 2004, 2004.

[33] K. Wada and T. Shibata, "Living With Seal Robots - Its Sociopsychological and Physiological Influences on the Elderly at a Care House," IEEE Trans. Robot., 23(5), pp. 972-980, 2007.

[34] R. S. Wilson et al., "Loneliness and Risk of Alzheimer Disease," *Arch Gen Psychiatry*, Vol. 64:234-240, Feb. 2007.

[35] S. Yohanan et al., "Affect-Driven Emotional Expression with the Haptic Creature," UIST '09, 2009.

[36] T. Yoshikai, M. Hayashi, A. Kadowaki, T. Goto and M. Inaba, "Design and Development of a Humanoid with Soft 3D-Deformable Sensor Flesh and Automatic Recoverable Mechanical Overload Protection Mechanism", in IROS 2009, 2009.

[37] H. Zhang, A. C. Berg, M. Maire and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition", in *Proc. CVPR*, 2006.