25th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
August 26-31, 2016. Columbia University, NY, USA

Touch Recognition and Learning from Demonstration (LfD) for Collaborative Human-Robot Firefighting Teams

Wallace Lawson, Keith Sullivan, Cody Narber, Esube Bekele, and Laura M. Hiatt

Abstract— **In Navy human firefighting teams, touch is used extensively to communicate among teammates. In noisy, chaotic, and visually challenging environments, such as among fires on Navy ships, this is the only reliable means of communication. The overarching goal of this work is to augment Navy firefighting teams with an autonomous robot serving as a nozzle operator; to accomplish this, the robot must understand the tactile gestures of its human teammates. Preliminary results recognizing touch gestures have indicated the potential of such an autonomous system to serve as a nozzle operator in human-centric firefighting scenarios.**

I. Introduction

Fire presents a significant threat to Naval vessels. If not dealt with quickly, fire can spread rapidly and result in a significant loss of life and property. For example, in 2012, a fire aboard the USS Miami resulted in $700 million dollars in damage and injured seven Navy personnel [16].

Navy firefighting teams typically consist of two people: a nozzle operator and a supervisor (see Figure 1). The nozzle operator actively works toward fire suppression, using the hose to engage the fire directly in front of them. The supervisor generates a higher level strategy on the most effective way to fight the fire in the given compartment. They do this by visually searching the environment for higher priority fires (e.g., fire near a person, fire near a gas supply). The supervisor can redirect the nozzle operator as needed to suppress higher priority fires. For safety, the nozzle operator always attends to the fire immediately in front of them, and does not visually search around for additional fires.

Effective communication in firefighting teams is vitally important. However, the chaotic nature of the environment can limit the ways that firefighting teams communicate. Although hand gestures are commonly used in human-human communication [8], [1], [5], they are used infrequently in firefighting due to extremely poor visibility. Instead, communications between the supervisor and nozzle operator are primarily through speech and tactile gestures. The supervisor touches the shoulders of the nozzle operator to communicate, for example, a command to turn, enter a compartment, leave a compartment, or reengage after the supervisor has returned from investigating another region of the environment. Speech is sometimes used to confirm the commands.

This paper works towards replacing the nozzle operator, perhaps the most dangerous situation for a navy firefirefighter, with a humanoid robot. One challenge with introducing a robot into the firefighting team is teaching the robot how to respond to human touch in order to understand the supervisor's commands. To study this problem, we use the humanoid robot, Octavia (see Figure 2), a mobile dexterous social (MDS) robot built to study human-robot interaction. Although not designed to fight fire, Octavia provides an excellent platform to study the underlying HRI principles of incorporating a robot into firefighting teams.

W. Lawson, K. Sullivan, C. Narber, E. Bekele, and L. Hiatt are with the Center for Applied Research in Artificial Intelligence, Naval Research Lab, Washington, DC 20375

Fig. 1. A Navy firefighting team engaging a fire in an enclosed compartment. The team consists of two individuals: the supervisor (behind), and the nozzle operator (in front).

Typically, tactile information is processed when a human makes contact with the robot, after which time the force sensing resistors are used to interpret the meaning of the touch. There are several problems with this approach. First, a significant amount of hardware is required for the force sensing resistors, and covering the entire surface of a robot with these sensors is infeasible. Second, interpretation in this manner can slow the robot's response to the human contact. Since firefighting requires fast responses between teammates, we use a different approach where we can anticipate touch before the human makes contact with the robot. Thus, rather than recognizing touch, we instead look for *visual touch gestures*, hand gestures that result in the supervisor touching the robot, where we anticipate the tactile information before a hand has made contact with the robot.

We make use of the Leap Motion [13] NIR sensor to recognize visual touch gestures. The Leap Motion sensor was designed to recognize hand gestures to control a computer or interact with a virtual environment, and is sensitive to a range of 0.1 m - 1 m which is sufficient to recognize visual touch gestures. We propose a novel approach to processing visual touch gestures with the Leap Motion. We demonstrate our approach to be both accurate and reliable, typically outperforming the Leap Motion SDK.

U.S. Government work not protected by
U.S. copyright

Fig. 2. The humanoid robot, Octavia, engaged in the firefighting task

Fig. 3. The Leap Motion Sensor mounted on Octavia's back. The position and angle of the sensor allows coverage of the most sensetive regions of the torso for firefighting interactions.

The rest of the paper is organized as follows. Section II briefly discusses related research in this area, while Section III provides technical details of our approach to tactile gesture recognition. Section IV then evaluates our approach to tactile gesture recognition, and Section V concludes the paper by discussing the implications of this work.

II. Related Work

This paper focuses on tactile communication between a human and robot in the context of firefighting. Most prior work in robotic firefighting has spanned the gamut from design [14] to sensing [11], [17]. Naghsh et al. [19] proposed to use a swarm of robots to fight a fire. In this case, the swarm is remotely controlled by a human operator. This limits the effectiveness of the team, as it is impossible for a human firefighter to communicate directly with the robots. Penders et al.[20] proposed to use a visor to provide situational awareness to the human firefighter by showing information on the swarm of robots. Again, this does not permit any direct interaction with the robots, and limits the potential for the robot to be a collaborative teammate. Closest to our work, Martinson et al. [15] developed an interaction system for robotic firefighting where the human directly communicates with robot through speech and visual hand gestures. This system performed well for communicating situational awareness to the robot prior to engaging the fire, but is not practical during firefighting operations since the human and robot must face each other.

While there is almost no prior work recognizing visual touch gestures in robotic firefighting, many researchers have investigated the use of hand gestures in other contexts. Much work has focused on hand gesture recongintion in a variety of contexts [4], [10], [23], [25], [2], but all these approaches assume that a continuous full view of the hand is available, and it is not clear that they will work in a situation where the hand is only partially visible as is often the case in visual touch gestures.

Some work on human robot interaction has focused on identifying different touch gestures. For example, Ji et al. [9] developed a system to distinguish between similar but subtly different types of touch (poke, grab, touching a large area). Wu et al. [24] similarly used a support vector machine to identify the orientation of the touch. A complete review of tactile HRI can be found in [3].

An earlier version of this paper proposed to use force sensitive resistors (FSR) to classifying different touch gestures [12]. While the approach worked well, the small size (5 cm by 5 cm) of each FSR requires a large number of sensors to cover all the areas of interest for firefighting interactions. The large number of sensors produces a significant volume of wires and other necessary hardware components which take up space on the robot and are a source of noise and failures. Additionally, the FSRs are require physical contact which results in a slow response from the robot.

In learning from demonstration, the term "tactile gestures" to refer to small "nudges" are used to communicate desired motor movements to a robot [18]. This is primarily used to teach movement information, and no attempt is made to understand the intent of the touch. Interpreting the meaning of touch is sometimes considered in the context of multi-touch displays [26]. Although there is some similarity, sensing with multi-touch displays is substantially different.

III. Features for Recognizing Visual Touch Gestures

To detect visual touch gestures, we use the Leap Motion Senor, a small (30mm x 76 mm x 13mm) NIR stereo camera which provides a wide field of view at short ranges. The Leap uses IR emitters to detect hand gestures for computer interaction (i.e., replace the standard mouse). While the Leap Motion comes with a proprietary SDK, in our experience most of the visual touch gestures were not detected due to the speed at which the hand is moving. Experiments are detailed below demonstrating this time delay. To cover the areas most sensitive for firefighting, the Leap Motion Sensor was attached to Octavia's back and angled upwards and away from the torso (Figure 3).

Our feature extraction procedure operates on the raw NIR images coming off the Leap Motion Sensor. Before detecting features, a preprocessing step enhances the edges and minimizes noise. Next, an adaptive thresholding algorithm detects

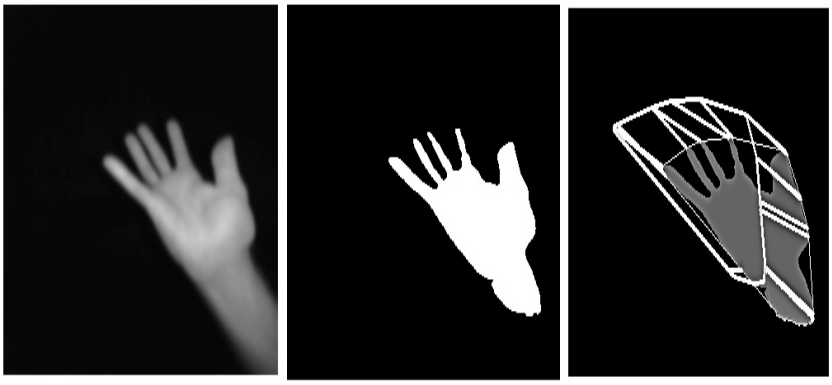

Fig. 4. Our feature extraction pipeline: the raw IR image from the Leap sensor, extracted hand blob, and motion vectors constructed from two successive frames.

objects that are close to the sensor (i.e., the hand/arm). The hand / arm is isolated with a convex hull and a feature vector is constructed based on the region inside this convex hull. The feature vector is a combination of static and dynamic features: static features relate to the static pose of the hand to include the location and the hand orientation, while the dynamic features provide information about how the hand is moving (i.e., speed and trajectory)

We construct a 13 element feature vector to describe visual touch gesture, using a combination of features used in other visual hand gesture recognition approaches[4], [10], [23], [25], [2]. The static features consist of the (1-2) X and Y image coordinates of the top-most hull point, a (3) depth estimate based on the number of pixels between the convex hulls in the two stereo images, and the (4) ratio of the convex hull perimeter to detected hand/arm blob perimeter. The dynamic features include the (5) change in depth between frames, and a (6-13) histogram of angles containing eight bins where each bin contains the magnitude of the movement vectors. Figure 4 shows an example of the raw NIR image, the associated hand blob, and the movement vectors. For each frame, we collect a feature vector for each hand: if one hand is not visible, then its feature vector is set to zero.

For each gesture sequence, we gather 20 frames of stereo image pairs and compute a feature vector of length 26, or 13 features per hand. The final feature vector is a concatenation of these 20 vectors, resulting a final feature vector of length 260. Finally, the feature vector is normalized to range [-1, 1].

IV. Complete Firefighting Robot

Visual touch gesture recognition fits into the larger system of a robot firefighter. The complete firefighting system consists of sub-systems to detect fires, detect visual touch gestures, and learn how to react to visual touch gestures. We use a learning from demonstration (LfD) system to train the robot to react to various visual touch gestures and fire locations.

A. Fire Detection

Fire detection is done using an IR camera. Due to noise in the raw IR images, we first process the image using a Gaussian blur smoothing filter followed by thresholding to isolate the brightest IR source, which is typically the fire. Successive dilation and erosion operations further smooth the images, and, then, individual IR blobs are extracted and convex hulls are fitted to these blobs. The center of the fire (i.e., the centroid of each convex hull) is used as an input to the LfD system.

B. Reaction to Visual Touch Gestures

Learning from demonstration (LfD) is an effective way to teach a robot to interact with humans [6], [7] since LfD is flexible to enough adapt to different teaching styles. We use the HiTAB (Hierarchical Training of Agent Behaviors) LfD system [22], [21] which learns behaviors represented as hierarchical finite state automaton (HFA) represented as a Moore machine: individual states correspond to hard-coded agent behaviors, or the states may themselves be another automaton. An HFA is constructed iteratively: starting with a behavior library consisting solely of atomic behaviors (e.g., turn, go forward), the demonstrator trains a slightly more complicated behavior, which is then saved to the behavior library. The now expanded behavior library is then used to train an even more complex behavior which is then saved to the library. This process continues until the desired behavior is trained.

The motivation behind HiTAB was to develop a LfD system which could rapidly train complex agent behaviors. HiTAB uses manual behavior decomposition as its primary method to reduce the size of the learning space, thus allowing rapid training in areas such as behavior based robotics, where samples are sparse and expensive to collect. HiTAB's behavior decomposition breaks complex joint finite-state automata into a hierarchy of simpler automata which are iteratively learned and then composed via scaffolding into more complex behaviors. This manual decomposition requires the experimenter to choose the automaton's general structure: HiTAB then determines the appropriate transitions.

To determine which behaviors to train, we observed videos of Navy firefighting teams during training to learn which behaviors teams typically perform. Primarily the nozzle operator focuses only straight ahead, engaging any observed fires, and does not search the environment. The supervisor is responsible for searching the environment for other, possibly higher priority, fires, and then using a combination of touch and speech direct the nozzle operator to engage those fires.

Based on these observations, we used HiTAB to train a behavior that searches for a visible fire, engages the fire, and will change which fire is engaged based on visual touch gestures from the human supervisor. The features available were the X-position of the fire (see Section IV-A), the classification of visual touch gestures as described earlier, and a *Done* flag used to signal parent automata that a child is complete. The basic behaviors were **Turn Left, Turn Right**, and **Idle**, which turn the torso or keep it stationary. Using these features and basic behaviors, we trained the following automata (Figure 5):

- TOUCH TURN LEFT: Wait in **Idle** until *Touch Left* is **True**, then **Turn Left** until *Touch Left* is **False**.
- TOUCH TURN RIGHT: Wait in **Idle** until *Touch Right* is **True**, then **Turn Right** until *Touch Right* is **False**.
- SEARCH FIRE: Using *Fire X-Position*, perform visual servoing to keep the fire in the center of the camera frame.
- SEARCH FIRE HRI: Adds an **Interested** expression to Octavia's face prior to performing visual servoing. This HFA acts as confirmation that Octavia has actually observed a fire.
- FIREFIGHTING: Combines speech and visual touch information to allow the human supervisor to refocus Octavia on a higher priority fire by turning Octavia whenever the supervisor touches the robot.

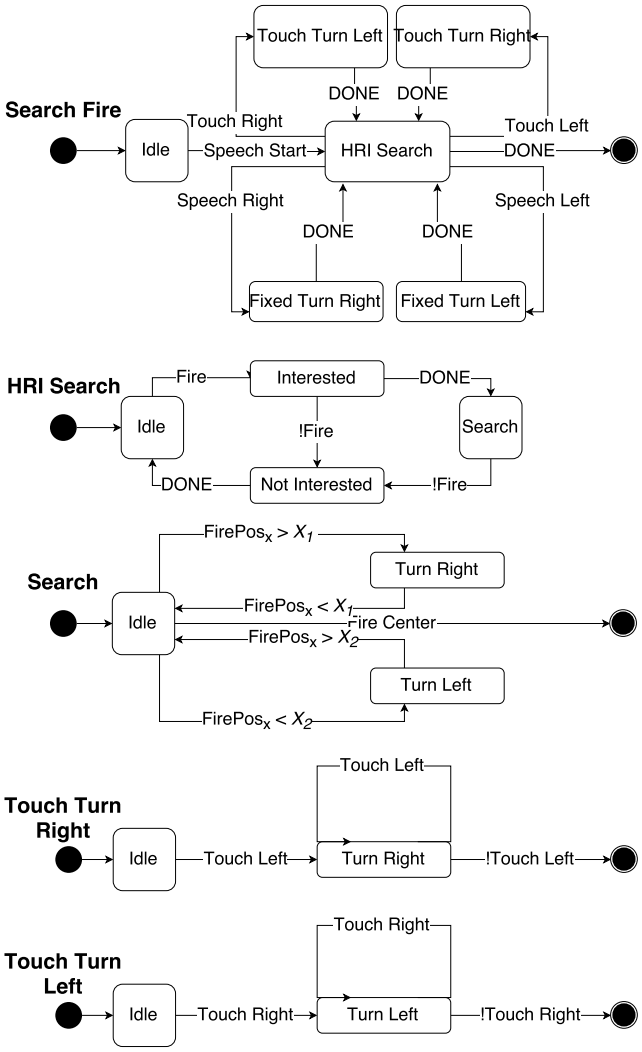

Fig. 5. Trained HFAs for firefighting.

V. EXPERIMENTAL RESULTS

In this section, we present our experimental results showing the accuracy of visual touch gesture recognition and to compare the accuracy of this system with the accuracy of the Leap Motion SDK.

A. Tactile Gesture Classification

We recognize visual touch gestures using classification approaches commonly used in hand gesture recognition. Specifically, we evaluate support vector machines (SVM) with a linear kernel, Random Forest (RF), Gradient Boosted Trees (GBT), and artificial neural network (ANN). The ANN is fully connected with a single hidden layer containing 60 nodes. Each node within the network uses a symmetric sigmoid activation function, and is trained using resilient back propagation. All the ensemble tree classifiers used 200 trees with 80% of samples and 70% of maximum features considered at each split of the training set. Moreover, we also added an AdaBoost (AdaB) and bagged (Bag) ensemble of 10 RF classifiers for complete comparative analysis.

To evaluate visual touch gesture recognition, we performed a user study to collect examples of hand gestures from individuals. The individuals interacted with Octavia, our humanoid robotics platform, providing a number of examples of each of the visual touch gestures. We present results using both 10-fold stratified sampling for cross-validation as well as a randomly shuffled 80%-20% split.

B. Accuracy of the Proposed System

To evaluate the effectiveness of the developed autonomous robotic firefighting system, we performed a user study involving 10 healthy adults to act as the human supervisor and collected data using the Leap Motion for visual touch gesture recognition. 10 healthy adults (9 Male and 1 Female) participated in the study performing 5 gestures each (of engage, turn left, turn right, pull, and miscellaneous). The miscellaneous class is added to avoid the case where any noisy hand movement would be classified in to one of the four important gestures for firefighting. Before each visual touch gesture, the user was asked to stand at a distance that represents a close working distance, around 1 m. The subjects were instructed to perform each gesture 10 times with variable ways of touching the robot including using dif-

ferent hands (for instance one handed touch vs. two handed touch), varying speed and individual ways of touching. The models need to learn these individual preferences of touching the robot for a more generalized recognition. 10 subjects performed 5 gestures with 10 repetitions of each gesture resulting in a total of 500 sample data points.

The results of this experiment are shown in Table I. The results show that statistically, there is little difference between the performance of the classifiers. Therefore, faster classifiers such as a linear SVM would be appropriate for our fast visual touch gesture recognition.

TABLE I

ACCURACY COMPARISONS OF THE CLASSIFICATION MODELS TRAINED WITH VARIOUS CROSS-VALIDATION METHODS

	10-fold	80%/20%
SVM	0.78 (+/- 0.08)	0.81 (+/- 0.03)
RF	0.80 (+/- 0.12)	0.88 (+/- 0.02)
GBT	0.80 (+/- 0.12)	0.87 (+/- 0.04)
ANN	0.80 (+- 0.01)	0.79 (+- 0.01)

Table II shows the confusion matrices of the best models in each of the two cross-validation methods compared. It is evident from these tables that the models were able to accurately recognize all the 4 meaningful gestures. However, the miscellaneous (Misc) gestures class was the worst classified in both instances. This is partially due to the way the Misc gestures were performed by each subject. Subjects were instructed to perform any hand gesture except the 4 gestures and there were large variability both across subjects and how similar some of the 'misc' gestures were to one of the 4 other meaningful classes of gestures.

TABLE II

CONFUSION MATRIX FOR THE TWO CROSS-VALIDATION METHODS WITH THE BEST PERFORMING MODELS. THE CONFUSION MATRIX SHOWS THE RESULTS FOR ADABOOSTED RF (TOP) AND BAGGED RF (BOTTOM)

	Engage	Turn Left	Turn Right	Pull	Misc
Engage	**77.8%**	0%	11.1%	1.1%	0%
Turn Left	0%	**100%**	0%	0%	0%
Turn Right	0%	0%	**90.9%**	9.1%	0%
Pull	0%	9.1%	0%	**90.9%**	0%
Misc	0%	10%	10%	20%	**60%**

	Engage	Turn Left	Turn Right	Pull	Misc
Engage	**77.8%**	0%	0%	11.1%	11.1%
Turn Left	0%	**100%**	0%	0%	0%
Turn Right	0%	0%	**100%**	0%	0%
Pull	0%	9.1%	0%	**90.9%**	0%
Misc	0%	0%	0%	30%	**70%**

C. Comparison with the Leap SDK

To compare the performance of our approach with the Leap SDK, we had several individuals perform each of the tactile gestures that we were interested in and tracked whether or not the leap SDK could detect a hand when compared to our technique. The results shown in Table III illustrate the need for our own feature extraction techniques. Although no recognition took pace by the Leap SDK, it is clear that many of the tactile gestures would not be detected at all.

One interesting note about the data is that turning right versus left is detected at greatly differing levels of accuracy (53.3% vs. 13.3%). This is likely due to several factors: where the users stood behind the robot, how the leap is mounted, and what the hand preference is of the users (for example all users here were right-handed).

TABLE III

PERCENTAGE OF TIMES WHERE THE LEAP SDK DETECTED A HAND

	Percentage
Engage	90.0%
Turn Left	53.3%
Turn Right	13.3%
Pull	23.1%
Overall	45.7%

VI. DISCUSSION AND CONCLUSION

Using autonomous robots as part of Navy firefighting teams presents a meaningful and difficult challenge. Here, we investigated how to enable the robot to communicate with its human teammates in the noisy and messy environment that firefighters operate in. In particular, we developed an approach to visual touch gesture recognition that allows the robot to quickly and effectively understand its human supervisor's tactile gestures, even when the provided data is sparse or incomplete.

To enable tactile gesture recognition for effective communication between the human supervisor and the robot nozzle operator, we have developed a novel way of extracting features and recognizing tactile touch in the partially visible and uncertain environment which performed much better than the SDK of Leap as shown by the validation results.

Early results indicated that the hard task of effectively equipping the robot with tactile recognition capabilities would facilitate the human supervisor and robot nozzle operator communication in a way similar to the human-centric firefighting.

In the future, we would like to continue to more deeply explore the rich nuances of tactile gestures. Gestures, for example, may have different levels of urgency (e.g., touching with extra speed and force), and the same gesture could be used to communicate different commands or intents depending on the context (e.g., single tap to get someone's attention, or to indicate spatial information). In future work, we propose to further extend our understanding of touch gestures to explore these nuances.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research (GT). The views and conclusions contained in this paper do not represent the official policies of the U.S. Navy. Keith Sullivan was supported by the Naval Research Laboratory

under a Karles Fellowship. We thank William Adams and Anthony Harrison for their assistance.

REFERENCES

[1] M. W. Alibali, M. Bassok, K. L. Olseth, S. Syc, and S. Goldin-Meadow. Gestures reveal mental models of discrete and continuous change. In *Proceedings of the seventeenth annual conference of the Cognitive Science Society*, pages 391–396, 1995.

[2] N. Anjum and A. Cavallaro. Multifeature object trajectory clustering for video analysis. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1555–1564, 2008.

[3] B. D. Argall and A. G. Billard. A survey of tactile humanrobot interactions. *Robotics and Autonomous Systems*, 58(10):1159 – 1176, 2010.

[4] M. Bhuyan, P. Bora, and D. Ghosh. Trajectory guided recognition of hand gestures having only global motions. *International Journal of Computer Science, Fall*, 2008.

[5] R. Breckinridge Church and S. Goldin-Meadow. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23(1):43–71, 1986.

[6] S. Chernova and A. L. Thomaz. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(3):1–121, 2014.

[7] M. J. Gielniak, C. K. Liu, and A. L. Thomaz. Generating human-like motion for robots. *The International Journal of Robotics Research*, 32(11):1275–1301, 2013.

[8] S. Goldin-Meadow. Gesture with speech and without it. *Duncan Cassell Levy*, pages 31–50, 2007.

[9] Z. Ji, F. Amirabdollahian, D. Polani, and K. Dautenhahn. Histogram based classification of tactile patterns on periodically distributed skin sensors for a humanoid robot. In *RO-MAN, 2011 IEEE*, pages 433–440, July 2011.

[10] Y.-G. Jiang, Q. Dai, X. Xue, W. Liu, and C.-W. Ngo. Trajectory-based modeling of human actions with motion reference points. In *Computer Vision–ECCV 2012*, pages 425–438. Springer, 2012.

[11] J.-H. Kim, J. W. Starr, and B. Y. Lattimer. Firefighting robot stereo infrared vision and radar sensor fusion for imaging through smoke. *Fire Technology*, 51(4):823–845, 2015.

[12] W. E. Lawson, K. Sullivan, and G. Trafton. Understanding touch gestures on a humanoid robot. In *2014 AAAI Fall Symposium Series*, 2014.

[13] LeapMotion. http://www.leapmotion.com/, June 2014.

[14] P. Liljeback, O. Stavdahl, and A. Beitnes. Snakefighter - development of a water hydraulic fire fighting snake robot. In *Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on*, pages 1–6, Dec 2006.

[15] E. Martinson, W. Lawson, S. Blisard, A. Harrison, and G. Trafton. Fighting fires with human robot teams. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[16] J. McDermott. Fire-stricken submarine uss miami is decommissioned, March 2014. [Online; posted 29-March-2014].

[17] B.-C. Min, E. T. Matson, A. Smith, and J. E. Dietz. Using directional antennas as sensors to assist fire-fighting robots in large scale fires. In *Sensors Applications Symposium (SAS), 2014 IEEE*, pages 360–365, Feb 2014.

[18] K. Mülling, J. Kober, O. Kroemer, and J. Peters. Learning to select and generalize striking movements in robot table tennis. *The International Journal of Robotics Research*, 32(3):263–279, 2013.

[19] A. Naghsh, J. Gancet, A. Tanoto, and C. Roast. Analysis and design of human-robot swarm interaction in firefighting. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 255–260, Aug 2008.

[20] J. Penders, L. Alboul, U. Witkowski, A. Naghsh, J. Saez-Pons, S. Herbrechtsmeier, and M. El-Habbal. A robot swarm assisting a human fire-fighter. *Advanced Robotics*, 25(1-2):93–117, 2011.

[21] K. Sullivan. *Hierarchical Multiagent Learning from Demonstration*. PhD thesis, George Mason University, 2015.

[22] K. Sullivan, S. Luke, and V. A. Ziparo. Hierarchical learning from demonstration on humanoid robots. *Proceedings of Humanoid Robots Learning from Human Interaction Workshop*, 38, 2010.

[23] T. Wenjun, W. Chengdong, Z. Shuying, and J. Li. Dynamic hand gesture recognition using motion trajectories and key frames. In *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, volume 3, pages 163–167. IEEE, 2010.

[24] H. Wu, H. Liu, and D. Liu. Two-dimensional direction recognition using uniaxial tactile arrays. *Sensors Journal, IEEE*, 13(12):4897–4903, Dec 2013.

[25] M.-H. Yang, N. Ahuja, and M. Tabb. Extraction of 2d motion trajectories and its application to hand gesture recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(8):1061–1074, 2002.

[26] Y. Yuan and K. Barner. An active shape model based tactile hand shape recognition with support vector machines. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 1611–1616, March 2006.