Abstract—Hand gesture recognition enables non-tactile interfaces for human-machine interactions. Cameras are currently powerful tools to recognize these gestures, however, use of cameras is constrained by privacy concerns and need for well-lit, line of sight implementation. In this study, we propose an alternate method to recognize gestures using a passive data-glove augmented with passive RFID tags. We envision passive tags-based gesture recognition will have applications in improving operator safety around machines, activity monitoring in factories and sign to speech recognition, etc. Low-level reader information (RSSI, Phase and Doppler frequency) can be used to capture changes to the tags in the environment, therefore generating enough information to infer gestures. In this paper, we present a technique to enable fast feature recognition using low-level reader data by correcting for inconsistencies in phase data due to frequency hopping. We experimented with four different classifiers on the low-level reader data and our Fully-Connected Neural Network (FCCN) classifier is able to learn gestures from tag-data with 98% accuracy.

I. INTRODUCTION

Recent developments in the Internet of Things (IoT) suggests a trend in developing advanced battery-powered sensors and smart devices with on-board radio and computing power. These devices are capable of acquiring rich environmental information seamlessly, however, battery replacements on these devices and higher device costs limit their pervasive implementation [1]. Billions of everyday objects still remain to be augmented with sensors to extend human interaction with machines and objects in the environment with inexpensive and scalable alternatives, which presents a huge potential within the IoT to use passive tags. Passive Radio Frequency Identification (RFID) tags provide such inexpensive, scalable and energy efficient way to gather abundant environmental information needed to generate contextual awareness [2]. We present a method for gesture recognition by augmenting everyday safety gloves with passive tags. A commercial off-the-shelf reader such as Impinj’s Revolution R420 can read low-level information (RSSI, Phase and Doppler Frequency) through the standard Gen 2 UHF RFID protocol [3]. This low-level information is affected by the tag’s dielectric background, location and objects in the environment – thereby has tag’s environmental information embedded. A single tag’s information can be affected by other parameters such as multi-path interference beyond our parameters of interest [4]. Instead, in this work, we use information from a combination of tags which is relatively robust compared to data from a single tag to identify features of interest. A passive data-glove will be low-cost and easy to use with potential applications in real-time operators’ safety and activity monitoring, and sign language to speech conversion as seen in Figure 1. Main contributions of this paper include:

- Method for gesture recognition using standard passive tags on everyday safety gloves
- Technique for correcting inconsistencies in phase data from successive channels in frequency hopping scheme increasing the usable sampling rate for sensing
- Classification models for low-level reader data in a multitag setup to recognize fine movements in the environment

The rest of the paper is organized as follows: Section II summaries the relevant state-of-the-art techniques for gesture recognition using camera/non-camera-based methods and human-activity recognition using passive RFID tags; Section III discusses our data-input approach and classification models to use low-level reader data; Section IV evaluates four different models and presents an outlook on future work and potential applications

II. RELATED WORK

A. Gesture recognition: camera and non-camera-based methods:

A recent review on gesture recognition methods is presented in this recent article by Cheok et al. [5]. Vision-based gesture recognition using cameras is a prominent technique due to it’s multi-functionality, stealth mode sensing, non-intrusive application. However, deploying such a system in everyday environments requires robust performance demanding good quality cameras or even expensive depth-cameras [6]. Continuous image or video analysis also creates significant data overhead that could be avoided with alternate options. Moreover, use of cameras in certain environments is currently a debatable topic involving users’ privacy concerns. Therefore, there is a potential for alternate methods for gesture recognition beyond

978-1-7281-1210-7/19/$31.00 ©2019 IEEE

978-1-7281-0589-5/19/$31.00 ©2019 IEEE
vision-based systems. Another widely explored method is using "data-gloves" that use gloves equipped with a variety of sensors. In one study, researchers used accelerometer sensors on each finger connected to a central hub to detect sign language gestures, however, the electronics increase the cost and make gloves cumbersome to wear [7]. In another study, Electromyogram (EMG) technique is used for hand gesture recognition using biosignals [8]. Even in this case, wearability of EMG electrodes in everyday use is a concern. Wi-Fi signals have also been used to detect fine grain gestures, but robustness and scalability of such system needs evaluation [9]. In other studies, haptic feedback and flexible sensors are also explored for developing gesture recognition methods [10], [11]. In this study, we propose a passive data-glove based method using RFID tags that leverages the benefit of low-cost and easy to use nature of tags.

B. Activity sensing with RFID tags:

Passive RFID tags are also explored previously for gesture recognition [12]. However, this method from the available literature uses a grid of tags as a monitoring environment. By monitoring how a moving hand interfered with the grid of tags, user’s hand motion is traced to the corresponding cells in the grid. Scaling such a grid-based method to variety of environments is practically difficult. In our work, we draw inspiration from other studies where everyday objects are instrumented with RFID tags for a variety of human activity recognition. Tags are used for in-home human-activity and household object usage monitoring [2], [13]. In another study, tags are used for creating battery-less interfaces to interact with other objects using gestures [14]. In our study, instead of attaching tags to the objects, our objective is to create a generic glove with tags that can be used to recognize a large set of gestures. Augmented reality and computer vision applications can also leverage information from tags to create immersive experiences for users by connecting even low-cost everyday objects such as cups, water bottles, etc. to the network in smart spaces [15].

III. METHOD

A. Pre-processing phase information:

Prototype in this study is a passive "data-glove" created by augmenting an everyday safety glove with five tags representing five fingers as in Figure 2. Any gestural changes will result in a change in RF phase and/or RSSI parameters due to change in the tags’ orientations. In this work, we use five different gestures as illustrated in Figure 3. By comparing these changes among five tags, gestures can be inferred. However, the RF phase reported directly by the reader shows discontinuity over time due to the reader’s frequency hopping scheme. FCC regulations require RFID readers in the ISM band to hop transmission frequencies between 50 channels in 902 - 928 MHz in order to minimize RF signal interference. Moreover, Impinj readers measure phase with an offset at each frequency making the measurements even more inconsistent. Reader also adds a 180 deg phase ambiguity, further increasing the noise in the measurements. To minimize these inconsistencies, phase values are corrected by removing the frequency dependent offset relative to a reference frequency. In this study, we used the lowest frequency (902.75 MHz) in Impinj’s frequency hopping scheme as the reference and all RF phase measurements are corrected to this reference frequency. Correction offset is calculated by averaging RF phase drift per channel for a still tag in 902 - 928 MHz range. According to [16], resulting phase of a received tag signal can be expressed as:

\[
\text{\( \text{Phase} = \text{Phase}_{\text{propagation}} + \text{Phase}_{\text{tag}} + \text{Phase}_{\text{reader_offset}} \)}
\]

where \( \text{Phase}_{\text{propagation}} \) is due to the signal path length, \( \text{Phase}_{\text{tag}} \) is due to tag’s backscattering properties and \( \text{Phase}_{\text{reader_offset}} \) is due to the offset in the corresponding frequency channel.

Figure 4 shows a 70 seconds snapshot of raw RF phase and corrected phase values along with data-points at the reference frequency (reference data points are graphically common to both raw phase and corrected phase lines since all other data points are corrected with reference to them). In contrast to the corrected phase data, raw phase data does not show a clear trend with time, however, for a still tag we expect to have more
or less constant phase with small variations. By analyzing the same data in frequency-domain, we can see a linear drift in RF phase across frequency channels (Figure 5). We observe this variation is approximately $2 \pi$ across 50 frequency channels. Assuming the offset is not channel specific, we can calculate the correction offset as $2 \pi / 49$.

Using this correction factor, raw RF phase of an $i^{th}$ reading is adjusted by the following equation:

$$\text{Corrected \ phase}_i = \text{Raw \ phase}_i + \frac{2 \pi}{49} \times \text{Channel}_i$$

Corrected RF phase data as shown in Figure 4 is stable with +/- 0.3 rad variation. We applied this phase correction technique on our phase data from the "data-glove" before feeding into the classifier.

B. Data input:

Our classification models use RSSI, RF phase and Doppler frequency from five tags corresponding to five fingers as the input. Experiments are conducted with an antenna and glove separation of 80 cm with reader in session 2 and transmitting at 30 dBm power. We use standard slurp library in python for data acquisition. We initially record base line data when the glove is in its initial position/gesture, which in this case is the "fist" gesture. Then RSSI, RF phase and RF Doppler frequency are measured relative to the base line data that helps in visualizing the trends in these parameters. Input to our classifier is a 15x1 vector of containing each of these three parameters from five tags. We use a time window of 0.5 seconds that contains around 4 measurements from each tag. Therefore, the input vector is created by averaging measured values within this time window. One can also increase the window time to increase the robustness of classification.
TABLE I: Structure details of different models

(a) CNN           (b) FCNN
Layer | Filters (size, #) | Layer | # of neurons
--- | --- | --- | ---
1 | (5, 64) | 1 | 64
2 | (5, 128) | 2 | 128
3 | (3, 64) | 3 | 32

(c) SVM

| Category | # of support vectors |
|--- | --- |
| Fist | 231 |
| 1-finger | 239 |
| 2-finger | 240 |
| 3-finger | 245 |
| Palm | 239 |

(d) RFC

| Parameter | Value |
|--- | --- |
| # of trees | 10 |
| Avg. # of nodes | 244.6 |
| Avg. depth | 14.1 |

C. Gesture classifier:

In this study, to classify hand gestures, we experiment with four different methods: Convolutional Neural Network (CNN), Fully-Connected Neural Network (FCNN), Support Vector Machine (SVM) and Random Forest Classifier (RFC). Among them, CNN and FCNN are deep learning approaches (CNN extracts local information whereas FCNN aggregates global information during forward propagation process); SVM and RFC are machine learning methods (SVM is intrinsically designed for binary classification whereas RFC is more suitable for multi-class classification problem). Few examples of applying these methods on sensors data in available literature can be found here [17], [18], [19], [20]. All four models essentially play the same role by taking as input an 1D vector of low-level reader parameters from five tags, and outputting a probabilistic distribution among five predefined gestures with the element of the highest score indicating the recognized hand gesture, as shown in Figure 6. Details of these methods are summarized in Table I. More specifically, the CNN model includes three 1D convolution layers, and the size of each 1D kernel and the total number of them in each layer are shown in Table I (a); The FCNN model also contains 3 fully-connected layers with the number of neurons per layer shown in Table I (b). After these model layers, a fully-connected output layer with the number of neurons equaling to the number of output classes is attached to each model separately to generate desired probabilistic distributions. If the number of neurons is too large, the model results in over-fitting and if the number of neurons is too low, the model results under-fitting. Therefore, the number of neurons is intuitively chosen considering the dimensions of input data to avoid over-fitting/under-fitting. We implement an one-vs-one multi-class SVM model to handle more than two classes, and radial basis function kernels are used to compute kernel matrix from the input data matrix. After the learning process, the number of support vectors for each category is shown in Table I (c). For RFC, we implement a 10-tree random forest, as shown in Table I (d), after training process, the average number of nodes and and the average depth per tree are 244.6 and 14.1, respectively are obtained.

TABLE II: The number of collected samples

| Gesture | Fist | 1-Finger | 2-Finger | 3-Finger | Palm |
|--- | --- | --- | --- | --- | --- |
| # | 1154 | 1191 | 1200 | 1224 | 1193 |

IV. EVALUATIONS

A. Visualizing RF phase change with gestures:

We applied the aforementioned phase correction technique on RF phase data from the “data-glove”. A snapshot of 70 second raw data from five tags is inconsistent due to frequency hopping and reader phase offsets as shown in the Figure 7. However, upon correcting for the offset, a consistent trend is visible as in Figure 8. Further, any gestural movements show features in the trend as highlighted by activity windows in Figure 9. Therefore, simple phase correction technique as this enables us to use RF phase measurements from all frequency channels without waiting for cycles of a particular frequency channel (every 10 seconds) – enabling smaller time-windows in using RFID low-level reader data for activity sensing. These results also help in visually correlating gestural changes with RF phase measurements and to build better classification models.

Fig. 7: Raw phase data from five tags corresponding to each finger on the glove from a still hand

B. Classifier Performance Comparison:

In this paper, we experiment with four different kinds of classification models, and we are primarily interested in evaluating their performance.

Datasets: To train and evaluate the proposed classifiers, we collect 5,962 samples in total covering all five predefined hand gestures. Details of collected samples are summarized in Table II. For samples within each gesture category, we randomly split them into training and testing sets by following a split ratio of 0.8/0.2.
TABLE III: The detailed gesture recognition evaluation

| Gesture | CNN Precision | FCNN | SVM | RFC |
|---------|---------------|------|-----|-----|
| Fist    | 0.99          | 1.00 | 0.94| 1.00|
| 1-Finger| 0.96          | 0.98 | 0.84| 0.99|
| 2-Finger| 0.94          | 0.98 | 0.75| 0.96|
| 3-Finger| 0.94          | 0.97 | 0.79| 0.97|
| Palm    | 1.00          | 1.00 | 1.00| 0.99|
| Avg.    | 0.97          | 0.99 | 0.86| 0.98|

Evaluation Metrics: To quantitatively evaluate the performance of different models, we adopt four metrics: recall, precision, f1-score and overall accuracy. Recall, precision and f1-score values are computed within each gesture category, and their averaged values can be used to access model performance in general. The overall accuracy accounts for the percentage of all the testing samples of different hand gestures that are correctly recognized.

TABLE IV: The overall gesture recognition evaluation

| Model | Recall | Precision | F1-score | Accuracy |
|-------|--------|-----------|----------|----------|
| CNN   | 0.97   | 0.99      | 0.98     | 0.98     |
| FCNN  | 0.99   | 0.98      | 0.98     | 0.98     |
| SVM   | 0.86   | 0.86      | 0.86     | 0.86     |
| RFC   | 0.98   | 0.98      | 0.98     | 0.98     |

Results: We show detailed evaluation results for each gesture category in Table III and summarize their averaged/overall performance in Table IV. From Table IV, we can easily observe that FCNN and RFC both achieve the best results in terms of precision, f1-score and accuracy, and FCNN shows the highest recall score. SVM performs worst in general, because it is intrinsically designed for binary classification, even though it can be extended for multi-class classification problems, its performance is still quite limited. We also note that the CNN model shows slightly worse performance than the FCNN model, we argue this may be caused by the fact that local information presented in input vector is not as strongly correlated as images, where CNN models typically show much better performance. Moreover, the input data size is relatively small, thus a fully-connected network structure is more suitable to explore all the correlated information to make decisions. Similar performance trends can also be observed...
in Table III across different hand gestures. Figure 10 shows a confusion matrix of predicted class against the true class for FCNN. While predicting gesture of 2-fingers, around 95% times the classifier predicted correctly and 5% times gesture 2-Fingers is confused with 3-Fingers. We argue that activities involving “middle” and “ring” fingers are confused with each other as we can easily test that one cannot easily lift their ring finger similar to index finger from the initial fist position. To further improve the accuracy and robustness of classification, we need to collect more data points in a range of background scenarios.

Operator’s safety while working around machinery is often a concern. Real-time monitoring of operator’s activities, machine’s state of operation and compliance with interlocks can vastly improve safety in industrial setting. This can be achieved to some extend by augmenting standard safety-gloves (in cases where operators already use safety-gloves) with passive tags. These passive data-gloves will be ease to use and affordable compared to existing sensor-rich gloves. As a continuation to this work, we plan to increase the number of recognizable gestures to cover broad range of gestures from the American Sign Language and build a generic gesture to speech conversion system.

V. CONCLUSION

Hand gesture recognition using everyday safety gloves augmented with passive RFID tags enable us to create low-cost “connected-gloves” for industrial activity and safety monitoring applications. We use low-level reader information (RSSI, Phase and Doppler frequency) to capture relative changes to the tags on the glove, thereby generating enough information to infer different gestures. Raw RF phase data from the reader shows discontinuity due to reader’s frequency hopping schemes. We correct this discontinuity by mapping RF phase data to a single reference frequency unlike previous studies that use longer sampling windows to monitor phase changes at a particular frequency of interest or that perform a frequency sweep within a sampling window. This cross-frequency phase correction technique enables us to use maximum read rate of the commercial RFID readers without waiting for RF phase measurements to be taken at a particular frequency. We experimented with CNN, FCNN, SVM and RFC classifiers to identify five types of hand gestures using the low-level channel data from five tags corresponding to five fingers. Our best models is capable of predicting gestures with 98% accuracy on the test data. As a continuation to this work, we plan to increase the number of recognizable gestures to cover the breadth of American Sign Language and build a gesture to speech conversion system based on our passive “data-glove”.

REFERENCES

[1] Mahyar Shirvanimoghaddam, Kamyar Shirvanimoghaddam, Mohamad Mahdi Abolhasani, Majid Farhangi, Vaid Zahari Barsari, Hangyue Liu, Mischa Dobler, and Minoo Naebe. Paving the path to a green and self-powered internet of things. arXiv preprint arXiv:1712.02277, 2017.
[2] Hanchuan Li, Chieh-yih Wan, Rahul C Shah, Alanson P Sample, and Shwetak N Patel. Idact: Towards unobtrusive recognition of user presence and daily activities. In 2019 IEEE International Conference on RFID (RFID), pages 1–8. IEEE, 2019.
[3] Impinj. Speedway revolution reader application note - low level user data support. In Technical Notes, pages 1–16. Impinj, 2012.
[4] M. Jo, H. Y. Yoon, S. Cha, and H. Choo. Mobile rfid tag detection influence factors and prediction of tag detectability. IEEE Sensors Journal, 9(2):112–119, Feb 2009.
[5] Ming Jin Cheok, Zaal Omar, and Mohamed Hisham Jaward. A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10(1):131–153, 2019.
[6] Juan Pablo Wachs, Mathias Kölsh, Helman Stern, and Yael Edan. Vision-based hand gesture applications. Communications of the ACM, 54(2):60–71, 2011.
[7] Jakub Galka, Mariusz Masior, Mateusz Zaborski, and Katarzyna Barczewska. Inertial motion sensing glove for sign language gesture acquisition and recognition. IEEE Sensors Journal, 16(16):6310–6316, 2016.
[8] Jonghwa Kim, Stephen Mastnik, and Elisabeth André. Emg-based hand gesture recognition for realtime biosignal interfacing. In Proceedings of the 13th international conference on Intelligent user interfaces, pages 30–39. ACM, 2008.
[9] Sheng Tan and Jie Yang. Wifinger: leveraging commodity wifi for fine-grained finger gesture recognition. In Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing, pages 201–210. ACM, 2016.
[10] Anand Asokan, Allan Joseph Pothen, and Raj Krishnan Vijayaraj. Armatrona wearable gesture recognition glove: For control of robotic devices in disaster management and human rehabilitation. In 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), pages 1–5. IEEE, 2016.
[11] Ying Huang, Le Gao, Yunong Zhao, Xiaohui Guo, Caixia Liu, and Ping Liu. Highly flexible fabric strain sensor based on graphene nanoplatelet-polyaniline nanocomposites for human gesture recognition. Journal of Applied Polymer Science, 134(39):45340, 2017.
[12] Parvin Asadzadeh, Lars Kulik, and Egemen Tanin. Gesture recognition using rfid technology. Personal and Ubiquitous Computing, 16(3):225–234, 2012.
[13] Michael Buettner, Richa Prasad, Matthis Philopose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In Proceedings of the 11th international conference on Ubiquitous computing, pages 51–60. ACM, 2009.
[14] Hanchuan Li, Eric Brockmeyer, Elizabeth J Carter, Josh Fromm, Scott E Hudson, Shwetak N Patel, and Alanson Sample. Paperid: A technique for drawing functional battery-free wireless interfaces on paper. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 5885–5896. ACM, 2016.
[15] Yongbin Sun, Sai Nithin R Kantareddy, Rahul Bhattacharyya, and Sanjay E Sarma. X-vision: An augmented vision tool with real-time sensing ability in tagged environments. In 2018 IEEE International Conference on RFID Technology & Application (RFID-TA), pages 1–6. IEEE, 2018.
[16] Pavel V Nikitin, Rene Martinez, Shashi Ramanurthy, Hunter Leland, Gary Spiess, and KVS Rao. Phase based spatial identification of uhf rfid tags. In 2010 IEEE International Conference on RFID (IEEE RFID 2010), pages 102–109. IEEE, 2010.
[17] Ming Zeng, Le T Nguyen, Bo Yu, Olo J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In 6th International Conference on Mobile Computing, Applications and Services, pages 197–205. IEEE, 2014.
[18] Tribeni Prasad Banerjee and Swagatam Das. Multi-sensor data fusion using support vector machine for motor fault detection. Information Sciences, 217:96–107, 2012.
[19] Luis Pablo Prieto, Kshitij Sharma, Łt Kidzinski, María Jesús Rodríguez-Triana, and Pierre Dillon. Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. Journal of computer assisted learning, 34(2):193–203, 2018.
[20] Charissa Ann Ronao and Sung-Bae Cho. Human activity recognition with smartphone sensors using deep learning neural networks. Expert systems with applications, 59:235–244, 2016.