

# Human Identification via Unsupervised Feature Learning from UWB Radar Data

Jie Yin<sup>1</sup>, Son N. Tran<sup>2</sup>, and Qing Zhang<sup>2</sup>

<sup>1</sup> Discipline of Business Analytics, The University of Sydney Business School,  
Sydney, Australia

`jie.yin@sydney.edu.au`,

<sup>2</sup> The Australian E-Health Research Centre, CSIRO, Brisbane, Australia  
{`son.tran, qing.zhang`}@csiro.au

**Abstract.** This paper presents an automated approach to automatically distinguishing the identity of multiple residents in smart homes. Without using any intrusive video surveillance devices or wearable tags, we achieve the goal of human identification through properly processing and analyzing the received signals from the ultra-wideband (UWB) radar installed in indoor environments. Because the UWB signals are very noisy and unstable, we employ unsupervised feature learning techniques to automatically learn local, discriminative features that can incorporate intra-class variations of the same identity, and yet reflect differences in distinguishing different human identity. The learned features are then used to train an SVM classifier and recognize the identity of residents. We demonstrate the effectiveness of our proposed solution through extensive experiments using real data collected in real-life situations. Our findings show that feature learning based on  $K$ -means clustering, coupled with whitening and pooling, achieves the highest accuracy, when only a limited quantity of training data is available. This shows that the proposed feature learning and classification framework combined with the UWB radar technology provides an effective solution to human identification in multi-residential smart homes.

**Keywords:** human identification, unsupervised feature learning, UWB, smart home

## 1 Introduction

The ability to recognize daily activities of residents is a core premise of smart homes for assisting with remote health monitoring and providing assistive interventions. For a smart home with only one resident, by deploying various sensors in the living space and assuming all sensor data is generated by this only person, a resident's daily activities can be accurately recognized through various types of machine learning models. In reality, however, homes are often occupied by more than one residents. As a result, activity recognition models designed for single living environments fail to yield satisfactory results from environmentally

deployed sensors, because of the absence of evidence with regards to which sensors are triggered by whom. Therefore, human identification is one of the most crucial problems faced by multi-residential smart homes to fully realize their functionality and potential.

Computer vision systems have been widely used to recognize human identity [14]. However, they have limited performance in poor visibility conditions (e.g., at night), and inevitably raise privacy concerns. This restricts them to be deployed in real-life smart homes that require to respect the privacy and comfort of residents under monitoring [12]. Another popular solution to human identification is the use of wearable devices that need to be carried by residents. In such wearable systems, each resident carries a tag which continuously advertises its unique ID through various types of wireless communication, such as Wi-Fi [6], Bluetooth [1], or RFID. These tags can be used as unique identifiers to help distinguish individuals. Each of these tags, however, is always assumed to be carried by a particular resident, and any of its presence is simply associated with that resident. These wearable systems require the residents to always remember and carry wearable devices all day, which forbids them to be widely accepted by senior communities, let alone by people with neurodegenerative diseases.

In this work, we achieve the goal of detecting the identity of residents in smart homes through properly processing and analyzing the received signals from UWB radar. UWB radar systems can be installed in indoor environments in a non-intrusive way, offering many advantages such as high-resolution ranging, low power cost, and strong resistance to narrowband interference [3, 7, 21]. They have abilities to detect the changes within the vicinity of a UWB radar when people pass by. However, due to the multi-path effect, the received UWB signals are very noisy; they are very sensitive to changes in the environment, as well as differing walking modes of people. Thus, the major challenge is to find robust features that are invariant enough to incorporate naturally occurring intra-class variations, for example, resulting from differing walking modes of people or changes in surrounding environments, but discriminative enough to distinguish between different classes. Another significant challenge is collecting large amounts of labeled data for learning, which is very difficult, labor-intensive, and sometimes even impossible when considering practical operation of smart homes. Thus, it is required to build models that are both accurate and lightweight to enable effective and efficient human identification in real-life smart homes.

To cope with these challenges, we propose an unsupervised feature learning and classification framework to recognize the identity of multiple residents in smart homes. This framework utilizes unsupervised feature learning based on  $K$ -means clustering to automatically learn a sparse representation from the UWB signals. Specifically, we consider UWB signals in a two-dimensional space and discover discriminative local features that capture useful patterns while filtering out redundant noisy information. This way, the original UWB signals are locally transformed into sparse representations that have desirable properties for distinguishing different human identity. The derived new features are then used to train an SVM classifier to recognize the identity of individuals.

We demonstrate the effectiveness of our proposed solution through extensive experiments using real UWB data collected from eight participants in real-life scenarios. We compare against several deep learning models, including sparse auto-encoders, sparse Restricted Boltzmann machines (RBMs), and convolutional neural networks (CNNs), and investigate the effect of several important factors (i.e., whitening, pooling, etc.) that affect their performance. We show that, despite its simplicity,  $K$ -means based feature learning achieves the highest accuracy, when only limited training data is available. This testifies its ability to discover discriminative patterns that are effective for human identification.

## 2 Related Work

The ultra-wideband technology has been widely used to identify targets in both military and civilian applications owing to its strong penetrability, high resolution, and anti-interference ability. A method for detecting human presence using UWB impulse-based radar in urban environments is proposed in [3]. Several other methods use UWB radar to detect and classify targets in foliage environments [7, 21]. These methods rely on manually extracted features, such as energy, maximum amplitude, or excess delay of received signal, to perform classification. In contrast, we consider UWB signals in a two-dimensional space and automatically learn discriminative features for distinguishing human identity.

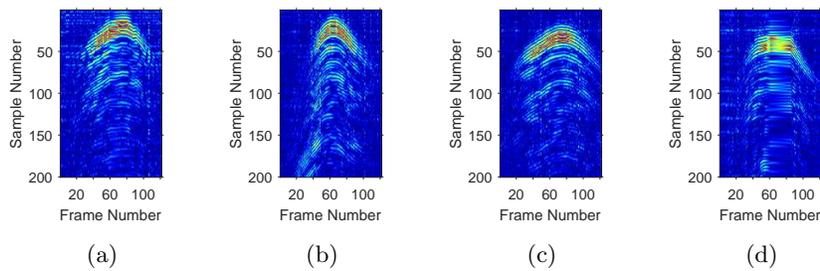
Much recent work in machine learning has focused on learning good feature representations from unlabeled input data to facilitate subsequent analytic tasks such as pattern recognition and classification. Deep architectures trained in an unsupervised manner have been proposed as an automatic method for extracting useful features. The focus has been on building different variants of deep neural networks with many hidden layers to learn multi-level representations, such as sparse auto-encoders [5], sparse RBM [8, 17], and CNNs [18]. Although these recently introduced algorithms often yield better classification results on benchmark datasets like MNIST [16] and CIFAR [13] for handwriting/object recognition, they must require careful selection of multiple hyper-parameters, such as learning rates, number of hidden nodes, etc., in search of higher performance. Unfortunately, how to tune these hyper-parameters is a non-trivial task; their values are often carefully chosen via cross-validation, thus dramatically increasing the amount of training data and running times. Moreover, they require large amounts of labeled data to achieve the current state-of-the-art results; their performance may dramatically degrade with insufficient training data.

In the computer vision area, extensive research has been dedicated to extracting higher-level image features to achieve better performance on object recognition or scene categorization [15]. It has been shown that the classification of natural images can be significantly improved using a multi-stage architecture of feature learning [11]. Among others,  $K$ -means clustering, as the unsupervised learning module in these feature learning pipelines, can lead to excellent results, often beating state-of-the-art systems [4]. In this work, we analyze the UWB signals which are essentially different from natural images that are widely studied

in computer vision; the UWB data is noisy, uncertain, and does not contain clear visual objects or edges, making it difficult to identify useful visual features. Thus, instead of working on the full-sized UWB data, we extract random patches and discover local patterns that are discriminative for recognizing human identity.

### 3 Characteristics of the UWB Data

Subject to reflection, refraction, diffraction, and even absorption by human body and the surrounding structures, UWB signal propagation suffers from severe multi-path effect; the impinging UWB electromagnetic wave scatters from different human body parts at different times with various amplitudes, depending on the distance to the body part, and the size and material of the reflecting part.



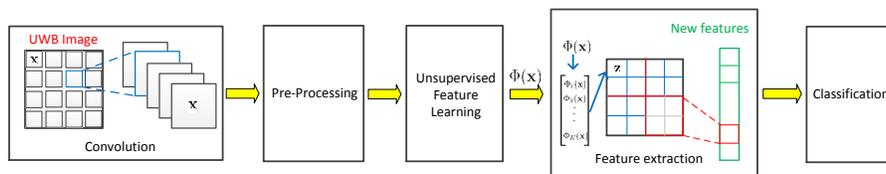
**Fig. 1.** Scattered UWB signals of two different subjects walking within the vicinity of a UWB radar. (a) and (b) indicate Subject A’s straight and diagonal walking. (c) and (d) indicate Subject B’s straight and random walking. It is evident that the local patterns of the UWB signals are quite different for different subjects, or the same subject but with differing walking modes.

Fig. 1 shows scattered UWB signals of two different subjects walking within the vicinity of a UWB radar, where brighter colors indicate closer distance between the UWB radar and the target. Because the two subjects have different height and body shape, when they pass by the UWB radar, the UWB signals reflected from their bodies have different strength; the signals scattered from a taller subject would be stronger as the signals travel shorter distance to and back from his reflecting body (e.g., head, shoulder). As can be seen, scattered UWB signals of Subject A (Fig. 1(a)) are noticeably different from those of Subject B (Fig. 1(c)) when they both pass by the UWB radar along a straight line.

Moreover, due to the multi-path effect, the reflected UWB signals might vary a lot even when the same subject passes by using different walking modes; for example, a person might walk by randomly or diagonally, instead of strictly following a straight line. As an example of UWB radar’s sensitivity to different walking modes, consider Fig 1(c) and 1(d) that exhibit different local patterns of UWB signals when Subject B takes a straight/random walk.

## 4 Unsupervised Feature Learning and Classification for Human Identification

To cope with the noisy characteristics of the UWB signals, the major challenge is to discover class models that are invariant enough to handle intra-class variations of the same subject, such as differing walking modes or gaits, and yet discriminative enough to distinguish between different subjects. Because there is a lack of prior knowledge about what features are potentially useful, we resort to automatically learning useful features from the unlabeled data and use them for classification. We extract random patches from the UWB signals and discover local patterns that are discriminative for distinguishing human identity.



**Fig. 2.** The workflow of unsupervised feature learning and classification

The workflow of unsupervised feature learning and classification for human identification is shown in Fig. 2. It consists of five major components: (1) convolutional module that generates random patches from the original UWB data, (2) pre-processing module that performs normalization and whitening on the input patches, (3) unsupervised feature learning module that learns a mapping function from an input patch to a new feature vector, (4) feature extraction module that transforms an original UWB data into a new representation using the learned mapping function, and (5) classification module that takes the new representation as input and trains a classifier for identity recognition.

### 4.1 Convolutional Module

Given a set of input UWB data  $U = \{u^{(1)}, \dots, u^{(n)}\}$ , we start by extracting random patches from unlabeled data, since we expect most useful, discriminative features are localized to a small region. Suppose each patch has dimension  $p$ -by- $p$ , where  $p$  is referred to as the patch size. Each  $p$ -by- $p$  patch can be represented as a vector  $x^{(i)} \in \mathbb{R}^N$  of pixel intensity values, with  $N = p \times p$ . This way we construct a dataset of  $m$  patches  $X = \{x^{(1)}, \dots, x^{(m)}\}$  that are randomly sampled from the input UWB data  $U$ .

### 4.2 Pre-Processing

Before applying any feature learning algorithm, it is useful to perform several pre-processing steps on the input patches  $x^{(i)}$ . The first step is normalization which

normalizes the brightness and contrast of the patches. For each patch  $x^{(i)}$ , we subtract out the mean of its elements and divide by the standard deviation, given by  $x^{(i)} = (\tilde{x}^{(i)} - \mu_{\tilde{x}^{(i)}})/\sigma_{\tilde{x}^{(i)}}$ , where  $\mu_{\tilde{x}^{(i)}}$  and  $\sigma_{\tilde{x}^{(i)}}$  are the mean and standard deviation of any unnormalized patch  $\tilde{x}^{(i)}$ . After normalization, we apply the ZCA whitening [10] on the entire dataset  $X$ . This process is commonly used in deep learning to remove correlations between nearby pixels. We will later empirically assess the usefulness of whitening on the UWB data for identity recognition.

### 4.3 Unsupervised Feature Learning

After pre-processing steps, we now apply an unsupervised learning algorithm on dataset  $X$  to extract discriminative features. Specifically, we aim to learn a function  $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^K$ , which maps an input vector  $x^{(i)} \in \mathbb{R}^N$  to a new  $K$ -dimensional feature vector  $z^{(i)} \in \mathbb{R}^K$ . In this work,  $K$ -means clustering is used as the unsupervised feature learning module. The classic  $K$ -means algorithm finds cluster centroids that minimize the distance between data points and the nearest centroids. In our context, the data points are randomly extracted patches and the centroids are the filters that will be used to newly encode the data. From this perspective,  $K$ -means algorithm learns to construct a dictionary  $\mathcal{D} \in \mathbb{R}^{N \times K}$  from the input vector  $x^{(i)} \in \mathbb{R}^N$  for  $i = 1, 2, \dots, m$ , such that the reconstruction error can be minimized. We use a modified version of  $K$ -means, similar to spherical  $K$ -means [9], that aims to find the dictionary  $\mathcal{D}$  according to:

$$\begin{aligned} \min_{\mathcal{D}, z} \quad & \sum_i \|\mathcal{D}z^{(i)} - x^{(i)}\|_2^2, \\ \text{s.t.} \quad & \|z^{(i)}\|_0 \leq 1, \forall i \quad \|\mathcal{D}^{(j)}\|_2 = 1, \forall j \end{aligned} \quad (1)$$

where  $z^{(i)} \in \mathbb{R}^K$  is a code vector associated with the input  $x^{(i)}$ , and  $\mathcal{D}^{(j)}$  is the  $j$ 'th column of the dictionary  $\mathcal{D}$ . Here, the objective is to find a dictionary  $\mathcal{D}$  and a new representation,  $z^{(i)}$ , of each input  $x^{(i)}$ , that can minimize the difference between  $x^{(i)}$  and its corresponding reconstruction  $\mathcal{D}z^{(i)}$ . The objective function is optimized under two constraints. The first constraint,  $\|z^{(i)}\|_0 \leq 1$ , indicates that each  $z^{(i)}$  is restricted to have at most one one-zero entry. The second constraint,  $\|\mathcal{D}^{(j)}\|_2 = 1$ , requires that each dictionary column has unit length. The formulation of objective function is very similar in spirit to learning coding schemes, such as sparse coding [20], which requires to solve a convex optimization problem and thus is difficult to scale up. In contrast,  $K$ -means has been proved very useful for learning features due to its efficiency and scalability [4].

This optimization problem (1) can be easily solved by alternatively optimizing  $\mathcal{D}$  and  $z^{(i)}$  as follows:

$$z_j^{(i)} = \begin{cases} \mathcal{D}^{(j)\top} x^{(i)} & \text{if } j = \arg \max_l |\mathcal{D}^{(l)\top} x^{(i)}|, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

$$\mathcal{D} = XZ^\top + \mathcal{D}, \quad (3)$$

$$\mathcal{D}^j = \frac{\mathcal{D}^{(j)}}{\|\mathcal{D}^{(j)}\|_2}, \quad (4)$$

where matrices  $X \in \mathbb{R}^{N \times m}$  and  $Z^{K \times m}$  have columns  $x^{(i)}$  and  $z^{(i)}$ , respectively.

Finally, the columns of the dictionary  $\mathcal{D}$ ,  $\mathcal{D}^{(j)}$ 's, are returned as  $K$  cluster centroids. Each patch  $x^{(i)}$  is mapped to a new  $K$ -dimensional vector  $z^{(i)} \in \mathbb{R}^K$ , with each element being the “distance” to the corresponding cluster centroid.

#### 4.4 Feature Extraction

After obtaining a new representation  $z^{(i)} \in \mathbb{R}^K$  for each patch  $x^{(i)}$ , we can build a new representation of any given UWB data. Given a  $w$ -by- $v$  UWB data, we define a  $(w-p+1)$ -by- $(v-p+1)$ -by- $K$  array of features by computing the representation  $z$  for every  $p$ -by- $p$  patch of the input data. To alleviate invariance to small distortion, we further reduce the dimensionality of the new representation via average pooling that combines local regions of  $z^{(ij)}$ 's using an average operation. Concretely, we split the  $z^{(ij)}$ 's into  $s$  spatial regions, and compute the average  $z^{(ij)}$ 's in each region. This results in a reduced  $K$ -dimensional representation for each pooling region, and a total of  $s \times K$  new features for each UWB data.

#### 4.5 Classification

Given these pooled feature vectors for a set of training UWB data and their class labels, we adopt a linear support vector machine (SVM) algorithm to train a classifier for human identification. When an unseen UWB data is observed, we first apply the same feature extraction method described above to obtain a new representation, and then apply the learned classifier to predict its class label, that is, the corresponding human identity.

## 5 Experiments

In this section, we validate the performance of our proposed solution using the UWB data collected from a real-world scenario. We first describe the set-up for data collection and pre-processing, and then discuss experimental results.

### 5.1 Data Collection and Pre-Processing

We used the system as shown in Fig. 3 to collect the UWB data in an indoor environment [19]. The system mounts a UWB transmitter and a UWB receiver on the top of the door frame at the entrance of a room. The UWB waveforms are propagated through the transmitter antenna, with a bandwidth of 3-6 GHz, into the detection zone of the radar. When people are passing the door, the UWB waveforms are scattered from human body and the surrounding objects, and are received at the receiver antenna via multiple paths. The UWB sensor is connected via USB to a PC to record the received signals.

For the task of human identification, eight subjects with different body shape and height participated in the experiments. The characteristics of these subjects

are listed in Table 1. To simulate how people pass the door in real life, each subject was requested to pass the door using four walking modes: straight walking, diagonal left-to-right walking, diagonal right-to-left walking, and random walking. For each walking mode, each subject continuously passed the door back and forth until he was asked to stop. Thus, we collected continuous UWB data sequences that contained multiple regions of interest that indicate the presence of participating subjects within the vicinity of the UWB radar.

Subject	Gender	Height (cm)
A	Male	170
B	Male	171
C	Male	168
D	Male	177
E	Male	169
F	Female	165
G	Male	160
H	Male	176

**Table 1.** Participating subjects



**Fig. 3.** Data collection set-up

For each UWB data sequence, we employed Canny edge detection [2] combined with density thresholding for efficiently localizing subjects in the sequence. The first step was to remove sparse signals to emphasize possible regions of interest where subjects might be detected. Second, we accumulated the densities along the time and frequency dimensions within a sliding window and used a threshold to find a center of a bounding box. With this we segmented regions of interest with dimension of  $200 \times 121$ , but due to the complexity in human’s trajectory the localization was not perfect. This is why treating each data sample as a set of local batches is needed in this work. In total, we obtained 768 samples of UWB data, with 96 samples per subject.

For evaluation, we randomly partitioned the data into three sets, each maintaining the same class distribution. We used the first two sets for training, and the third set was held out for testing. We compare  $K$ -means clustering with two other feature learning algorithms, sparse auto-encoders and sparse RBMs. For each feature learning algorithm, we learned features from either raw data or whitened data, and trained a linear SVM classifier that is tested on the holdout data. We evaluate classification performance using accuracy.

For all feature learning algorithms, the number of pooling regions  $s$  is set to 4, and the number of random patches is set to 100,000. For sparse auto-encoders and sparse RBMs, the regularization term of sparsity penalty is set to 0.001.

## 5.2 Effect of Whitening and Patch Size

We first performed experiments to test the performance of all the algorithms with respect to different patch sizes both with whitening and without whitening. We expected that larger patch sizes would allow us to discover more complex

features that cover a larger region of the original data. On the other hand, this increases the dimensionality of patches to be processed and may require to use more data for learning or to learn more features. In this experiment, we tested the patch size of 4, 6, 8, 10, 12, and 14 pixels, and set the number of features to 100. The classification results are shown in Table 2.

**Table 2.** Effect of whitening and patch size

Patch size	4	6	8	10	12	14
Sparse RBM	43.04	47.47	53.16	58.23	60.75	46.83
Sparse RBM (white)	47.47	60.13	67.72	77.22	72.15	71.52
Sparse auto-encoder	40.51	48.73	53.80	55.06	55.06	50.82
Sparse auto-encoder (white)	58.23	68.35	<b>77.22</b>	70.25	48.73	53.80
<i>K</i> -means	77.85	77.85	73.41	75.94	74.68	74.68
<i>K</i> -means (white)	<b>79.11</b>	<b>80.34</b>	76.58	<b>77.85</b>	<b>76.58</b>	<b>77.31</b>

Table 2 clearly shows the benefit of whitening: all algorithms generally achieve higher classification accuracy by performing whitening on the random patches. This confirms that whitening is a crucial pre-process for all the feature learning algorithms. We can see that, *K*-means with whitening achieves the highest accuracy in most cases, except that with the 8 pixel patch size, sparse auto-encoder performs slightly better than *K*-mean clustering. This is particularly notable because *K*-means is easy to implement and requires much less turning, unlike sparse auto-encoders and sparse RBMs which require us to carefully choose several hyper-parameters to guarantee reasonable results. Despite of its simplicity, the feature vectors produced by *K*-means constitute a powerful sparse and localized basis for distinguishing different human identity. Overall, the 6 pixel patch size works best for *K*-means that achieves an accuracy of 80.34%.

### 5.3 Effect of Number of Feature Bases

Since the number of feature bases  $K$  is an important parameter, we carried out experiments by varying the value of  $K$  from 50, 80, 100, 200, to 400. The value of  $K$  is equivalent to the number of centroids, for *K*-means clustering, and the number of hidden units, for sparse auto-encoders/RBMs. These experiments used the 6 pixel patch size and whitening for all algorithms.

Fig. 4 shows classification accuracy of all algorithms with respect to different values of  $K$ . We can see that, at the very beginning, all algorithms generally achieve higher performance by learning more features. This is expected because with more features learned, all algorithms have greater representative power, resulting in higher classification accuracy. However, at the later stage, learning more features decreases the accuracy, because the learned features are less distinguishable. Overall, *K*-means clustering achieves the best accuracy when the number of centroids is 100. Sparse RBMs and sparse auto-encoders performs best when the number of hidden units is 80 and 200, respectively. On the other hand, learning more features indeed increases the computational load and may

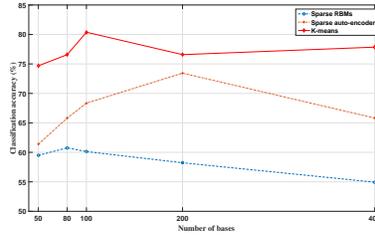


Fig. 4. Effect of number of feature bases

require more data for training. Thus, carefully choosing an appropriate number of feature bases can help achieve a good trade-off between classification accuracy and computational efficiency.

#### 5.4 Final Classification Results

We have shown through previous studies that whitening, a 6 pixel patch size, and 100 feature bases work best in general across all feature learning algorithms. Using these parameters, we ran our full pipeline on the training set, trained an SVM classifier, and evaluated on the test set. We also compare against two other baselines: (1) Raw features + SVM, which trains an SVM classifier on the vectorized UWB data, (2) CNN, which is one state-of-the-art deep learning model trained on the original UWB data. Our final results are reported in Table 3.

Table 3. Classification accuracy

Algorithm	Test Accuracy (%)
Raw features + SVM	28.22
CNN	58.07
Sparse RBM	59.49
Sparse auto-encoder	68.35
K-means	<b>80.34</b>

From Table 3, we can see that the three feature learning algorithms, sparse-auto-encoders, sparse RBMs and  $K$ -means clustering outperform the other two baselines that directly work on the original UWB data. This confirms that local patterns learned from input patches have more discriminative power than original features. Although CNN has demonstrated to render high performance on many image classification tasks, in the case that we only a limited number of training data, it produces unsatisfactory results for human identification. Again,  $K$ -means clustering achieves the highest performance with 80.34% accuracy.

Table 4 shows classification confusion matrix for identifying eight participants, with features learned by  $K$ -means clustering. As can be seen from the

**Table 4.** Confusion matrix with features learned by  $K$ -means clustering

truth / predict	A	B	C	D	E	F	G	H	Accuracy (%)
A	<u>17</u>	0	2	0	1	0	0	0	85.00
B	0	<u>20</u>	0	0	2	0	0	2	83.33
C	2	0	<u>16</u>	0	0	2	0	0	80.00
D	0	0	0	<u>13</u>	0	1	0	0	92.86
E	1	1	0	2	<u>12</u>	0	1	0	70.59
F	0	0	2	5	2	<u>13</u>	1	1	56.52
G	0	0	0	0	1	1	<u>17</u>	0	89.47
H	0	0	0	0	2	1	0	<u>17</u>	85.00
Overall									80.34

table, most of the errors occur when distinguishing participants with close height, for example, participants A (170 cm), B (171 cm), C (168 cm), and E (169 cm). Participants D (177 cm) and G (160 cm) are classified correctly with high accuracy of 92.86% and 89.47%, because they are the tallest and shortest among eight participants. Unexpectedly, the classifier sometimes has difficulty in distinguishing participant F (165 cm, female) from other male participants, yielding 56.52% accuracy only. This probably because she has similar body gait with some other males when walking within the detection zone of the UWB radar.

## 6 Conclusion

This paper proposed an automatic approach to human identification in multi-residential smart homes. We argued that, due to privacy or burden concerns, previous approaches to human identification that rely on video surveillance or wearable devices are not suitable as a practical solution in real-life smart homes. Instead, we achieved the objective of human identification through properly processing and analyzing the received signals from the UWB radar. We investigated the use of unsupervised feature learning techniques to automatically learn local, discriminative features for human identification. We evaluated our proposed solution through extensive experiments using real data collected from eight participants. Our results showed that  $K$ -means based feature learning, coupled with whitening and pooling, yields the best performance. This suggested that, while more complex algorithms like sparse auto-encoders or sparse RBMs may have greater representative power, they may not always be the best choice, given that in practice there may exist only a limited number of training data. In the future, we plan to implement our proposed solution in a real-time system to recognize human identity on a larger dataset and test its real-time effectiveness and efficiency.

## Acknowledgment

The authors would like to thank Chad Hargrave and Ghassem Mokhtari for constructive discussions. The UWB data was collected with ethics approval from CSIRO Health and Medical Research Ethics Committee, LR 12/2017.

## References

1. A. Alhamoud, A. A. Nair, C. Gottron, D. Bhnstedt, and R. Steinmetz. Presence detection, identification and tracking in smart homes utilizing bluetooth enabled smartphones. In *Proc. of the LCN Workshops*, pages 784–789, 2014.
2. J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, June 1986.
3. S. H. Chang, N. Mitsumoto, and J. W. Burdick. An algorithm for UWB radar-based human detection. In *Proc. of the IEEE Radar Conference*, 2009.
4. A. Coates, Honglak, and A. Y. Ng. An analysis of single-layer networks in unsupervised feature learning. In *Proc. of AISTATS*, pages 215–223, 2011.
5. I. Goodfellow, H. Lee, Q. V. Le, A. Saxe, and A. Y. Ng. Measuring invariances in deep networks. In *NIPS*, pages 646–654, 2009.
6. Q. Hao, F. Hu, and Y. Xiao. Multiple human tracking and identification with wireless distributed pyroelectric sensor systems. *IEEE Systems Journal*, 3(4):428–439, 2009.
7. X. He and T. Jiang. Target identification in foliage environment using uwb radar with hybrid wavelet-ICA and SVM method. *Physical Communication*, 13:197–204, 2014.
8. G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.
9. K. Hornik, I. Feinerer, M. Kober, and C. Buchta. Spherical  $k$ -means clustering. *Journal of Statistical Software*, 50(10):1–22, 2012.
10. A. Hyvarinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, 2000.
11. K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *Proc. of ICCV*, pages 2146–2153, 2009.
12. P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. Exploring privacy concerns about personal sensing. In *Proc. of PerCom*, pages 176–183, 2009.
13. A. Krizhevsky. Learning multiple layers of features from tiny images. Master’s thesis, Department of Computer Science, University of Toronto, 2009.
14. J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for easyliving. In *Proc. of the 3rd IEEE International Workshop on Visual Surveillance*, pages 3–10, 2000.
15. S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. of CVPR*, pages 2169–2178, 2006.
16. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 86(11):2278–2324, 1998.
17. H. Lee, C. Ekanadham, and A. Y. Ng. Sparse deep belief net model for visual area V2. In *NIPS*, pages 873–880, 2008.
18. H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proc. of ICML*, pages 609–616, 2009.
19. G. Mokhtari, Q. Zhang, C. Hargrave, and J. Ralston. Non-wearable uwb sensor for human identification in smart home. *IEEE Sensors*, 99, 2017.
20. B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37(23):3311–3325, 1997.
21. S. Zhai and T. Jiang. Target detection and classification by measuring and processing bistatic UWB radar signal. *Measurement*, 47:547–557, 2014.