Wi-Fi Based Handwritten Signature Verification Using a Triplet Network

Young-Woong Kwon, Jooyoung Kim, and Kar-Ann Toh
School of Electrical and Electronic Engineering
Yonsei University
Seoul, Korea

ABSTRACT

Attributed to the omnipresence of the radio signals for communications, sensing and recognition utilizing the Wi-Fi signals has significant advantage in terms of accessibility over conventional sensing means such as the camera. However, utilizing the raw Wi-Fi signals to capture in-air handwritten signatures for identity verification is yet a challenging task. In this paper, we propose a system for identity verification based on the handwritten signature signals captured by the Wi-Fi Channel State Information (CSI). A triplet network is adopted to learn the correlation between the captured signals and the user identities. To facilitate a fast converging loss model, a kernel and the range space learning is initially adopted for mining the triplet inputs. Subsequently, the triplet network is trained on a ConvNet structure based on the mined triplet inputs. Our experiments on a Wi-Fi dataset collected in-house show encouraging verification accuracy with faster training loss convergence comparing with that of the baseline triplet network and the Siamese network.

CCS Concepts

•Security and privacy → Biometrics; •Computing methodologies → Neural networks;

Keywords

Wi-Fi signature signal; in-air handwritten signature verification; the Kernel and the Range space projection learning; triplet network

1. INTRODUCTION

Over recent years, several behavioral biometric traits have attracted attention in view of their rigid physical body independence. Among these behavioral biometrics, the signature-based user authentication [1, 2, 3] has attracted considerable interest with the development of in-air signature recognition systems [4, 5, 6, 7]. With the help of sensors such as the depth camera [4, 5] or a mobile sensor [6], the in-air signature recognition system has lower the spatial constraint in the process of signature acquisition comparing with contact-based authentication systems [8, 9, 10].

Recently, the commercial Wi-Fi device has been adopted for in-air signature authentication due to its easy accessible property [7]. Based on the distortion of the Wi-Fi CSI signal according to the user's gestures, the in-air signature recognition system showed reasonable user verification performance [7]. More recently, some studies attempted to implement the deep learning algorithms in Wi-Fi signal-based user authentication systems to improve the verification performance [11, 12].

In this paper, we utilize a deep triplet network for identity verification based on the Wi-Fi CSI signature signal. To achieve not only the desired verification accuracy but also a fast training speed, we adopt the kernel and the range (KAR) space learning [13, 14, 15, 16] in order to mine the distinctive triplet inputs. Subsequently, the triplet network which utilizes the ConvNet [17] structure as a feature extractor is trained based on the $L2$ distance comparison.

The main contributions of our work can be summarized as follows:

- Proposal of a system for identity verification based on the Wi-Fi handwritten signature signals using a deep triplet network.
- Adopted the kernel and the range (KAR) space learning in order to mine the distinctive triplet inputs which boosted the convergence speed of the training loss in the triplet network.
- Provision of an experimental study using a Wi-Fi handwritten signature dataset which was collected in-house based on 50 subjects.

The paper is organized as follows: related works including the triplet network and KAR space learning are introduced in Section 2 for immediate reference. Our proposed method is discussed in Section 3. Section 4 describes our experimental results and analysis. Some concluding remarks are given in Section 5.

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

2. RELATED WORKS

2.1 Triplet network

The triplet network is considered a metric learning based model which aims to learn useful representations by means of distance comparison [18]. It is often seen in person re-identification [19, 20, 21, 22] where the individual identities are matched based on discriminative image features. The main difference between person identification and re-identification is that the later is a more challenging task where images of the same person taken from different cameras or under different occasions are to be associated. In order to address our challenging Wi-Fi based verification task, we adopt a triplet network which optimizes the input data space so that data points with the same identity are closer to each other than those with different identities [23]. The triplet network receives triplet pairs of data as its input. These data triplets are constructed based on a combination of the input data. Since not all triplet samples contribute to the desired classification, recent attention has been paid to the choice of relevant input pairs for training. In order to optimize the training process which utilizes only some parts of the triplet pairs, several researches [20, 21, 22] generated triplets from a small number of classes (persons) in each iteration. In [24], a triplet mining process was implemented to speed up the training convergence. They utilized a large mini-batch at each training iteration and selected the triplets based on network training instead of random sampling. However, this strategy needed a few thousands of exemplar mini-batches in every training iteration for triplet pairs selection. This results in a heavy computational load in training. In order to make use of the small sample training size as well as to speed up the triplet training process, we adopt the kernel and the range space method for the learning.

2.2 Kernel and the range space learning

Generally, the multilayer feedforward neural networks is trained based on the gradient descent method via backpropagation [25]. However, setting the learning parameters such as the learning rate and the learning momentum is a time consuming task.

Recently, a gradient-free learning framework based on the kernel and the range (KAR) space manipulation has been developed for multilayer network learning [13, 14, 15, 16]. The learning method is grounded on linear algebra without needing any training iteration.

Given m training samples. Let $\mathbf{X} \in \mathbb{R}^{m\times(n+1)}$ denotes the training data set and $\mathbf{G} \in \mathbb{R}^{m\times n}$ denotes the network output. Then the multilayer neural network can be written as follows:

$$\mathbf{G} = \sigma\left(\left[\mathbf{1}, \sigma\left(\ldots\left[\mathbf{1}, \sigma\left(\left[\mathbf{1}, \sigma\left(\mathbf{X}\mathbf{W}_1\right)\right]\mathbf{W}_2\right)\right]\ldots\mathbf{W}_{(n-1)}\right)\right]\mathbf{W}_n\right), \tag{1}$$

where $\mathbf{w}_1 \in \mathbb{R}^{(n+1)\times h_1}, \mathbf{w}_2 \in \mathbb{R}^{(h_1+1)\times h_2}, \ldots, \mathbf{w}_n \in \mathbb{R}^{(h_{(n-1)}+1)\times n}$ are the network weight matrices, $\mathbf{1} = [1,\ldots,1]^T \in \mathbb{R}^{m\times 1}$ is the bias vector, and $\sigma(.)$ is the activation function. By adopting an one-hot encoded target $\mathbf{Y} \in \mathbb{R}^{m\times n}$, training of the weight matrices $\mathbf{W}_i$ using the KAR space method [16] can be computed as follows:

$$\mathbf{W}_i = \left[\mathbf{1}, \sigma\left(\ldots\left[\mathbf{1}, \sigma\left(\left[\mathbf{1}, \sigma\left(\mathbf{X}\mathbf{W}_1\right)\right]\mathbf{W}_2\right)\right]\ldots\mathbf{W}_{(i-1)}\right)\right]^{\dagger}\sigma^{-1}\left(\mathbf{Y}\right), \quad i = 1,\ldots,n. \tag{2}$$

3. PROPOSED SYSTEM

In this section, we propose an identity verification system based on the Wi-Fi in-air handwritten signature (which will be called Wi-Fi signature hereafter) using the triplet network [18]. Fig.1 shows an overview of the proposed system utilizing the kernel and the range (KAR) space learning [14, 16] for mining the triplet inputs. Essentially, the KAR space projection learning is utilized to learn the triplet input data by mining those hard positive and the hard negative samples from each of the given anchor sample (see item (a) in Fig. 1). The hard positive and the hard negative samples refer to positive and negative class samples which are likely to be misclassified by the network. Subsequently, the ConvNet structure in the triplet network (see item (b) in Fig. 1) is trained with the mined triplet data based on a triplet loss function using the $L2$ distance comparison (see item (c) in Fig. 1). The following subsections describe the details of the triplet mining using KAR space learning and the triplet network.

3.1 Triplet mining using the kernel and the range space learning

The network receives a triplet set of data as its inputs. These triplet data consist of the reference data (will be called anchor samples hereafter) and the corresponding positive class data (same class with that of the anchor) and the negative class data (different class from that of the anchor). The goal of the triplet network is to position the feature vectors with appropriate separation space by putting the positive samples close to the anchor sample while keeping the negative sample away from the anchor sample.

According to [24], it is important to select the hard positive samples and the hard negative samples with reference to the given anchor sample for fast loss convergence when training the triplet network. A hard positive sample is defined as a sample whose distance to the anchor sample is large (which is most likely to be misclassified as a negative sample). On the other hand, a hard negative sample is defined as a sample whose distance to the anchor sample is small (which is most likely to be misclassified as a positive sample). However, from the raw data, there is no information regarding whether a sample is considered hard positive or hard negative before we train the network.

In this work, we propose to adopt the kernel and the range (KAR) space learning (see Section 2.2 for details) as a pretraining network to mine the hard positive/negative samples from the given anchor sample. Since the KAR space learning has no iterative learning process, we can mine the triplet samples without using the time consuming backpropagation training process.

By training the network with the single shot KAR space learning, we can map the $L2$ distance among the samples by using the output vector of the KAR space network. Given a set of training data which is packed in matrix $\mathbf{X}$, the network output can be written as:

$$f\left(\mathbf{X}\right) = \sigma\left(\left[\mathbf{1}, \sigma\left(\ldots\left[\mathbf{1}, \sigma\left(\left[\mathbf{1}, \sigma\left(\mathbf{X}\cdot\mathbf{W}_1\right)\right]\mathbf{W}_2\right)\right]\ldots\mathbf{W}_{(n-1)}\right)\right]\mathbf{W}_n\right). \tag{3}$$

After training based on the KAR space projection, the given anchor sample $\mathbf{x}_{anc}$ can be used as the reference to decide whether the network output of a sample is far away from this anchor sample. In other words, a hard positive sample $\mathbf{x}_{pos}$ and a hard negative sample $\mathbf{x}_{neg}$ with respect to the

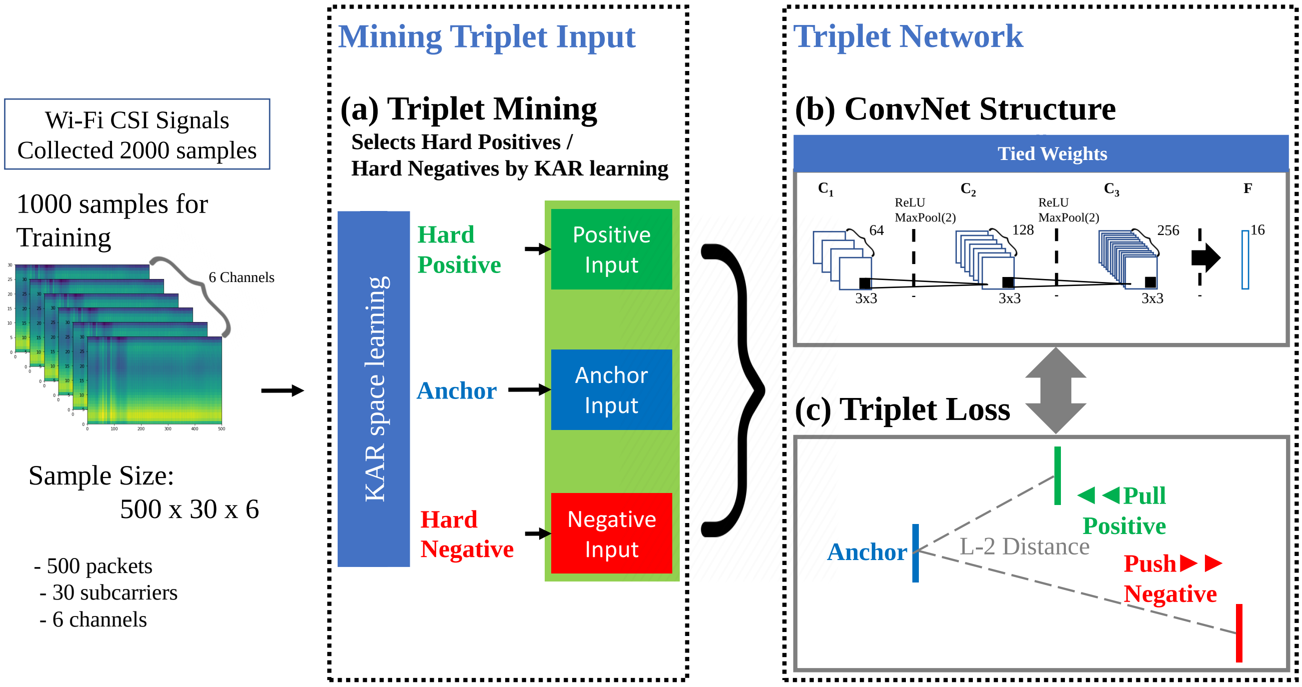


Figure 1: An overview of the proposed system.

anchor sample $\mathbf{x}_{anc}$ can each be determined based on:

$$\|f(\mathbf{x}_{anc}) - f(\mathbf{x}_{pos})\|_2^2 \geq \mathrm{t}_{pos}, \tag{4}$$

$$\|f(\mathbf{x}_{anc}) - f(\mathbf{x}_{neg})\|_2^2 \leq \mathrm{t}_{neg}, \tag{5}$$

where t_{pos} and t_{neg} respectively denote the thresholds that determine whether a sample is hard positive or hard negative. Since the hardest samples are likely to be outliers which can degrade the training process of the triplet network, we empirically set the t_{pos} at 75 percentile of the $L2$ distance and t_{neg} at 25 percentile of the $L2$ distance. The final set of $\mathbf{x}_{pos}$ and $\mathbf{x}_{neg}$ samples is randomly selected based on equations (4) and (5).

3.2 The ConvNet structure

The next step is to design a feature extractor which converts the triplet input data into feature vectors. In this work, we utilize the ConvNet structure [17] as a feature extractor since the three dimensional data format of our preprocessed input signal can be regarded as an image data format with multiple channels.

Our ConvNet design (item (b) in Fig 1) consists of three convolutional layers and one fully-connected layer. The number of convolutional filters to be trained in each layer is empirically chosen as $\{64, 128, 256\}$, with fixed filter size of 3×3, each of stride 1. The Rectified Linear (ReLU) activation function and the Max-pooling layer are applied between each convolutional layer. Subsequently, the extracted features from the last convolutional layer are flattened into a vector before feeding into the fully-connected network. The output vectors from the fully-connected layer are finally transformed using the sigmoid function followed by a $L2$ normalization.

3.3 The triplet loss

The triplet loss function was first seen in [18] for training the triplet network. For the i^{th} network input $\{\mathbf{x}_{anc,i}, \mathbf{x}_{pos,i}, \mathbf{x}_{neg,i}\}$, an anchor sample $\mathbf{x}_{anc,i}$ is randomly selected among the training data set. The positive and negative samples (respectively $\mathbf{x}_{pos,i}$ and $\mathbf{x}_{neg,i}$) are then determined and being selected for training the ConvNet based on the feature vectors $\{\mathbf{v}_{anc,i}, \mathbf{v}_{pos,i}, \mathbf{v}_{neg,i}\}$, for i=1,...,N. The triplet loss function is formulated based on a summation of the difference between the positive distance (the $L2$ distance between the anchor vector and the positive vector) and the negative distance (the $L2$ distance between the anchor vector and the negative vector) as follows:

$$loss = \sum_{i}^{N} max\left(\left[\|\mathbf{v}_{anc,i} - \mathbf{v}_{pos,i}\|_2^2 - \|\mathbf{v}_{anc,i} - \mathbf{v}_{neg,i}\|_2^2 + \alpha\right], 0\right), \tag{6}$$

where N denotes the size of the mini-batch, $\|.\|_2^2$ denotes the $L2$ distance and α denotes the preset margin. The ConvNet structure using equation (6) is trained to maximize the gap between the positive distance and the negative distance which should be larger than the margin α.

4. EXPERIMENTS

4.1 Dataset

In order to evaluate the verification performance of the proposed system, the Wi-Fi CSI signature dataset [7] is utilized in our experiments. The Wi-Fi CSI signature dataset consists of 2000 Wi-Fi CSI signature signals (4 directions $\times$ 10 samples $\times$ 50 identities) with sample size $500 \times 30 \times 6$. We

Table 1: The network structure of KAR space learning.

Layer	Size	Activation
Input	500×30×6	
Fully-Connected 1	1×1×1024	$\sigma = tan^{-1}$
Fully-Connected 2	1×1×16	$\sigma = tan^{-1}$
Output	1×1×50	

utilize only the absolute value from each complex CSI signal in our experiments.

4.2 Experimental settings

4.2.1 Performance evaluation:

The proposed system is evaluated under two cases: Case I on comparison between the proposed system and other handcraft or deep learning-based methods based on the verification error rate. Case II on comparing between the proposed system and the deep learning-based methods using the receiver operating characteristic (ROC) curve and training loss curve. For both cases, Wi-Fi signal measurements based on a single-orientation and 4-orientations of the user are recorded. For Case I, existing handcraft methods such as the least squares error estimation (LSE) [26], the principal components analysis (PCA) [27] with LSE, the support vector machine (SVM) [28] with different kernel functions, the total error rate minimization which adopted the reduced multivariate polynomial model as basis function (TER-RM2) [29], the deep learning-based Siamese network [30], and the baseline triplet network [18] are included for performance benchmarking. The Siamese network and the baseline triplet network utilized the same ConvNet structures with the proposed system and only differed in their input data style and loss function.
The verification performance of the proposed system and the compared methods are evaluated in terms of the Equal Error Rate (EER, %) which are taken from averaging the results of five runs of two-fold cross-validation tests. Due to the memory constraint caused by the large data size, the deep learning-based methods utilized randomly sampled 9,500 negative pairs in the validation stage, which is the same number as the number of positive pairs.

4.2.2 Network Structure and Parameter Settings:

The multilayer feedforward network structure of KAR space learning is specified in Table 1. The weights of each layer have been initialized with values of uniform distribution in [0, 1). We used $\sigma = tan^{-1}$ as the activation function following [15].
For the proposed system and the deep learning-based methods, we utilized the same ConvNet structure as specified in Table 2. We trained the network starting with a learning rate of 0.00005 and a mini-batch size of 32. We optimized the loss by the Adam optimizer with $L2$ penalty of 0.0002 except for the output layer. The output layer was regularized using an $L2$ penalty of 0.0001. We initialized all network weights in the convolutional layers with normal distribution of zero-mean and standard deviation 0.01. The biases were also initialized with a normal distribution of 0.5 mean and standard deviation 0.01. For the triplet networks, the hyper-parameter regulating triplet loss is empirically set at

Table 2: The structure of ConvNet model. For the convolution layer, the kernel is specified as (m×m) sized filter × (# of filters) / (# of stride). For the max-pooling layer, (p×p) sized pooling windows / # of stride. The input sizes are denoted as rows × cols × # of filters.

Layer	Activation	Kernel/Stride	Input Size
Conv 1	ReLU	(3×3)×64/1	500×30×6
MaxPool 1		(2×2)/1	500×30×64
Conv 2	ReLU	(3×3)×128/1	250×15×64
MaxPool 2		(2×2)/1	250×15×128
Conv 3	ReLU	(3×3)×256/1	125×8×128
MaxPool 3		(2×2)/1	125×8×256
Fully-Connected	Sigmoid	16	63×4×256
$L2$ Norm			1×1×16
Concat			1×1×16

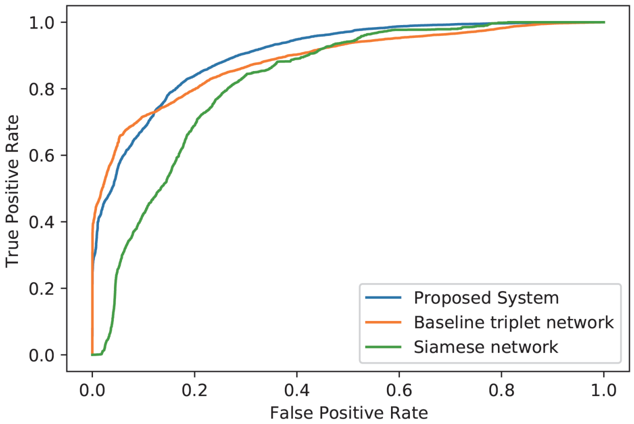
0.1. The training epochs were set at 1,500 for all three deep learning-based algorithms. For the linear methods such as LSE, SVM and TER, the input signals were resized to 500 × 30 by averaging along the subcarrier axes due to limitation of hardware memory. For the PCA-LSE, the input dimension was reduced to 40.

4.3 Results and discussion

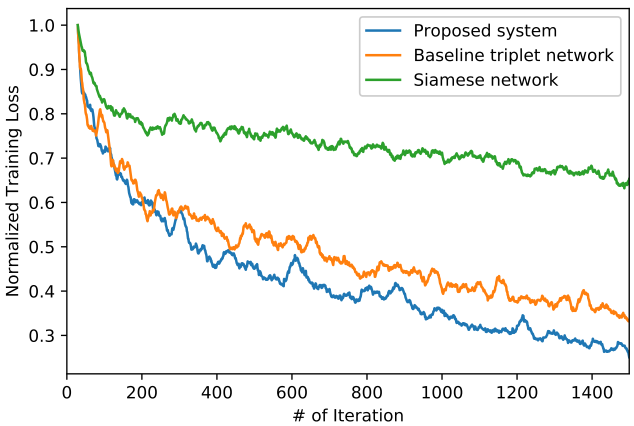
Case I: Table 3 shows the average of EER performance from five runs of two-fold cross-validation tests under the optimal parameter setting. For 1-orientation inputs, the SVM with Linear kernel shows the best verification accuracy of 1.55% EER. However, For 4-orientation inputs, all three deep learning-based methods show better performance than hand-craft methods. This is, perhaps, due to the utilization of the original size (500×30×6) of the input data in the training stage for the deep learning based methods. The test EER performance of the proposed system is 19.35%. The baseline triplet network without input mining showed slightly worse performance of 20.34% EER. The Siamese network showed the worst verification performance of 23.53% EER. The Siamese network is also a metric learning system, but differs from our system in that it receives two inputs and uses the contrastive loss function for training.
Case II: Fig. 2 (a) shows the ROC curves of the three compared deep learning-based methods. As shown in the figure, the proposed system clearly shows the largest Area Under Curve (AUC) among three compared methods. In Fig.2 (b) the training loss trends are plotted with respect to along the number of training iterations. Here, the proposed system shows the fastest training loss convergence followed by the baseline triplet network. In the figure, the y-axes of triplet network based methods and the Siamese network are normalized into $[0, 1]$ since each loss function has different starting value. According to these observations, it can be concluded that the triplet input mining with KAR space learning improves not only the verification performance but also the training loss convergence speed.

Table 3: Performance in terms of EER (%) averaged from five runs of two-fold cross-validation tests.

Methology	Equal Error Rate (%)		Parameter setting
	1-orientation	4-orientation	
LSE	35.40	48.44	
PCA-LSE	16.87	30.79	Reduced dimension=40
SVM(Linear)	1.55	28.23	c=1
SVM(RBF)	3.97	24.31	c=1, γ=0.01/3000
TER-RM2	13.29	35.84	M=1,τ=η=0.5
Siamese network	16.86	23.53	lr=0.00005
Baseline triplet network	5.45	20.34	lr=0.00005, α=0.1
Proposed system	**5.09**	**19.35**	**lr=0.00005, α=0.1**



(a) ROC Curves



(b) Normalized training loss trends

Figure 2: (a) Receiver Operating Characteristic (ROC) Curves and (b) Normalized training loss trends

5. CONCLUSION

In this paper, we proposed a system for identity verification based on the handwritten signature signals captured by the Wi-Fi Channel State Information (CSI). A triplet network adopting the ConvNet structure was used to learn the Wi-Fi information. In order to facilitate a fast training convergence, the inputs were first mined by a recent kernel and range space learning for samples which were likely to be misclassified. Subsequently, the network was trained with the mined triplet inputs based on the $L2$ distance comparison. Our experiments on an in-house Wi-Fi handwritten signature dataset showed not only an encouraging verification accuracy but also a faster training loss convergence compared with the baseline triplet network and the Siamese network.

6. ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2018R1D1A1A09081956).

7. REFERENCES

[1] Maged M.M. Fahmy. Online handwritten signature verification system based on DWT features extraction and neural network classification. *Ain Shams Engineering Journal*, 1(1):59–70, 2010.

[2] Ahmad Sanmorino and Setiadi Yazid. A survey for handwritten signature verification. In *the 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering*, pages 54–57. IEEE, 2012.

[3] Javier Galbally, Moises Diaz-Cabrera, Miguel A Ferrer, Marta Gomez-Barrero, Aythami Morales, and Julian Fierrez. On-line signature recognition through the combination of real dynamic data and synthetically generated static data. *Pattern Recognition*, 48(9):2921–2934, 2015.

[4] Je-Hyoung Jeon, Beom-Seok Oh, and Kar-Ann Toh. A system for hand gesture based signature recognition. In *the 12th International Conference on Control Automation Robotics & Vision (ICARCV)*, pages 171–175. IEEE, 2012.

[5] Jameel Malik, Ahmed Elhayek, Sheraz Ahmed, Faisal Shafait, Muhammad Malik, and Didier Stricker. 3DAirSig: A framework for enabling in-air signatures using a multi-modal depth sensor. *Sensors*, 18(11):3872, 2018.

[6] Hamed Ketabdar, Peyman Moghadam, Babak Naderi, and Mehran Roshandel. Magnetic signatures in air for

mobile devices. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion*, pages 185–188. ACM, 2012.

[7] Han-Cheol Moon, Se-In Jang, Kangrok Oh, and Kar-Ann Toh. An in-air signature verification system using Wi-Fi signals. In *Proceedings of the 4th International Conference on Biomedical and Bioinformatics Engineering*, pages 133–138. ACM, 2017.

[8] Yong Zhu, Tieniu Tan, and Yunhong Wang. Biometric personal identification based on handwriting. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 2, pages 797–800. IEEE, 2000.

[9] Anil K Jain, Friederike D Griess, and Scott D Connell. On-line signature verification. *Pattern recognition*, 35(12):2963–2972, 2002.

[10] Enric Sesa-Nogueras, Marcos Faundez-Zanuy, and Jiří Mekyska. An information analysis of in-air and on-surface trajectories in online handwriting. *Cognitive Computation*, 4(2):195–205, 2012.

[11] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, page 5. ACM, 2017.

[12] Akarsh Pokkunuru, Kalvik Jakkala, Arupjyoti Bhuyan, Pu Wang, and Zhi Sun. NeuralWave: Gait-based user identification through commodity WiFi and deep learning. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 758–765. IEEE, 2018.

[13] Kar-Ann Toh. Kernel and range approach to analytic network learning. *International Journal of Networked and Distributed Computing*, 7(1):20–28, December 2018.

[14] Kar-Ann Toh. Learning from the kernel and the range space. In *the IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, pages 1–6. IEEE, 2018.

[15] Kar-Ann Toh. Analytic network learning. *arXiv preprint arXiv:1811.08227*, November, 2018.

[16] Kar-Ann Toh, Zhiping Lin, Zhengguo Li, Beomseok Oh, and Lei Sun. Gradient-free learning based on the kernel and the range space. *arXiv preprint arXiv:1810.11581*, 2018.

[17] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[18] Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In *International Workshop on Similarity-Based Pattern Recognition*, pages 84–92. Springer, 2015.

[19] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: a deep quadruplet network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2017.

[20] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1335–1344, 2016.

[21] Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition*, 48(10):2993–3003, 2015.

[22] Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, and Lei Zhang. Joint learning of single-image and cross-image representations for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1288–1296, 2016.

[23] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[24] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.

[25] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[26] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[27] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[28] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[29] Kar-Ann Toh and How-Lung Eng. Between classification-error approximation and weighted least-squares learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):658–669, 2008.

[30] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015.