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, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
{herokwon, harrykim, katoh}@yonsei.ac.kr

Abstract. 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). To enable a fast loss convergence, a kernel and the range space learning is initially adopted for refining the triplet inputs by mining the distinctive inputs. Subsequently, the triplet network is trained on a ConvNet structure using the mined triplet inputs. Our experiments on an in-house Wi-Fi handwritten signature signal dataset show encouraging verification accuracy with faster training loss convergence comparing with the baseline triplet network and the Siamese network.

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 attracted attention in view of their rigid physical body independence [1]. Among these behavioral biometrics, the signature-based user authentication [15, 5] has attracted considerable interest with the development of in-air signature recognition systems [9, 17, 12]. With the help of sensors such as the depth camera [12] or a mobile sensor [9], the in-air signature recognition system has lower the spatial constraint in the process of signature acquisition comparing with contact-based authentication systems.

Recently, the commercial Wi-Fi device has been adopted for in-air signature authentication due to its easy accessible property [13]. 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 [13]. More recently, some studies attempted to implement the deep learning algorithms in Wi-Fi signal-based user authentication systems to improve the verification performance [18, 14].

In this paper, we utilize the 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 [20–22, 24] in order to mine the distinctive triplet inputs.

Subsequently, the triplet network which utilizes the ConvNet structure as a feature extractor is trained based on the L-2 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 boost the convergence speed of the training loss in the triplet network.
- Provision of an experimental study using an in-house Wi-Fi handwritten signature dataset collected from 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 are discussed in Section 3. Section 4 describes our experimental results and analysis. Some concluding remarks are given in Section 5.

2 Related works

2.1 Triplet network

The triplet network is considered a metric learning based model [26] which aims to learn useful representations by means of distance comparison [8]. It is often seen in person re-identification [2–4, 16, 25] 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 from the same camera in different occasions are to be associated. In this re-identification problem, to distinguish among the classes is a challenging task since the features are relatively weak compared with the background features. To address this problem, we adopt a triplet network which optimizes the embedding data space so that data points with the same identity are closer to each other than those with different identities [7].

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 [16] 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 [3, 4, 25] generated triplet from a small number of classes (persons) in each iteration. Recently, [16] adopted a triplet mining process to speed up the training convergence. They utilized a large mini-batch at each training iteration and selected the triplet based on the training network rather than 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.

2.2 Kernel and the range space learning [20–22, 24]

Generally, the multilayer feedforward neural networks is trained based on the gradient descent method via backpropagation [6]. However, setting the learning parameters such as the learning rate or 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 [21, 24]. The learning method is grounded on linear algebra with neither learning parameters nor iteration is needed in training.

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 outputs. Then the multilayer neural network structure can be written in linear equation form as follows:

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

where $\mathbf{W}_1 \in \mathbb{R}^{(n+1)\times h_1}$,$\mathbf{W}_2 \in \mathbb{R}^{(h_1+1)\times h_2}$,$\ldots$, $\mathbf{W}_i \in \mathbb{R}^{(h_{(i-1)}+1)\times n}$,$\mathbf{1} = [1, \ldots, 1]^T \in \mathbb{R}^{m\times 1}$ 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 [24] can be computed as follows:

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

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 [8]. Fig.1 shows an overview of the proposed system utilizing the kernel and the range (KAR) space learning [21, 24] for mining the triplet input. Essentially, the KAR space projection learning is utilized to learn the triplet input data by mining the hard positive and the hard negative samples from each 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 L-2 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 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 negative

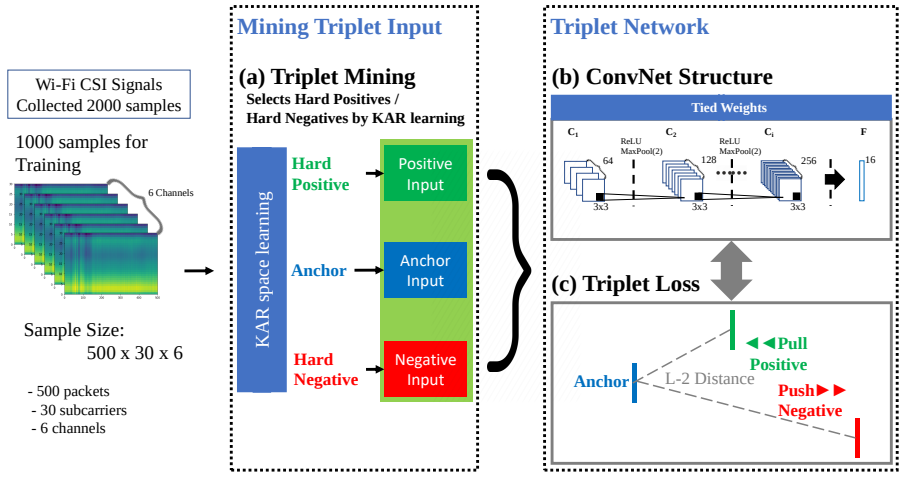


Fig. 1. An overview of the proposed system.

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 and keeping the negative sample away from the anchor sample.

According to [16], 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 about 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 L-2 distance between every samples by using the output vector of the KAR space network. **For training dataset $\mathbf{X}$, the trained network output is given by:**

$$f(\mathbf{X}) = \sigma\left(\left[\mathbf{1}, \sigma\left(\ldots\left[\mathbf{1}, \sigma\left(\left[\mathbf{1}, \sigma\left(\mathbf{X}\cdot\mathbf{W}_1\right)\right]\mathbf{W}_2\right)\right]\ldots\mathbf{W}_{(i-1)}\right)\right]\mathbf{W}_i\right), \quad (3)$$

For anchor sample $\mathbf{x}_{anc}$ which is randomly selected from training dataset $\mathbf{X}$, the output of the KAR learning makes it possible to mea-

sure the distance between each sample. A hard positive sample $\mathbf{x}_{pos}$ and a hard negative sample $\mathbf{x}_{neg}$ with respect to the anchor sample $\mathbf{x}_{anc}$ can each be respectively 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 L-2 distance and t_{neg} at 25 percentile of the L-2 distance. The final set of $\mathbf{x}_{pos}$ and $\mathbf{x}_{neg}$ samples is randomly selected from those samples threshold by (4) and (5).

3.2 ConvNet structure

The next step is to design a feature extractor which converts the input triplet data into feature vectors. In this work, we utilize the ConvNet structure [11] 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 structure (item (b) in Fig 1) for the network 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 L-2 normalization.

3.3 The triplet loss

The triplet loss function was first seen in [8] for training the triplet network. **For selected anchor input sample $\mathbf{x}_{anc,i}$, the i_{th} triplet input $\{\mathbf{x}_{anc,i}, \mathbf{x}_{pos,i}, \mathbf{x}_{neg,i}\}$ is generated by grouping positive input sample $\mathbf{x}_{pos,i}$ and the negative input sample $\mathbf{x}_{neg,i}$. The generated triplet input is fed into ConvNet structure and make feature vectors $\{\mathbf{v}_{anc,i}, \mathbf{v}_{pos,i}, \mathbf{v}_{neg,i}\}$.** The triplet loss function is formulated based on a summation of the difference between the positive distance (the L-2 distance between the anchor vector and the positive vector) and the negative distance (the L-2 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 L-2 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 [13] with the user located at a single position 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 500×30×6. We utilize only the absolute value from each complex CSI signal in our experiments.

4.2 Experimental settings

Performance evaluation: The proposed system is evaluated under two cases: i) case I on comparison between the proposed system and other handcraft or deep learning-based methods based on the verification accuracy, and case II on detailed comparison between the proposed system and the deep learning-based methods using the receiver operating characteristic (ROC) curve and training loss curve. For Case I, existing handcraft methods such as the least squares error estimation (LSE), the principal components analysis (PCA) [22] with LSE, the support vector machine (SVM) with different kernel functions and a total error rate minimization which adopted the reduced multivariate polynomial model as basis function (TER-RM2) [19, 23], the deep learning-based Siamese network [10] and the baseline triplet network [8] 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 other 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.

Network Structure and Parameter Settings: The multilayer feedforward network structure of KAR space learning is specified in Table 1. With input data, size of 500×30×6, we set two network layers where the size of the each layer is 1024 and 16, respectively. Each layer is initialized with uniform distribution over [0, 1). We used $\sigma = tan^{-1}$ as an activation function following [22].

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	

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 L-2 penalty of 0.0002 except for the output layer. The output layer was regularized using an L-2 penalty of 0.0001. We initialized all network weights in the convolutional layers with normal distribution of zero-mean and standard deviation of 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 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.

Table 2. The structure of ConvNet model. For the convolution layer, 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
L-2 Norm			1×1×16
Concat			1×1×16

4.3 Results and discussion

Case I: Table 3 shows the best average of EER performance from five runs of two-fold cross-validation test and the parameter condition. Among the handcraft methods, the SVM with RBF kernel showed the best verification accuracy

of 24.31% EER followed by the SVM with Linear kernel function. However, all three deep learning-based methods showed better performance than the SVM with RBF kernel function since the deep learning-based methods could utilize the original size (500×30×6) of the input data in the training stage. Among the deep learning-based methods, the Siamese network is also a metric learning system, but it differs from our system in that it receives two inputs and uses the contrastive loss function for training. The best average of test EER performance was obtained from the proposed system with 19.35% EER. 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. Case II: Fig. 2(a) shows the ROC curve of three deep learning-based methods.

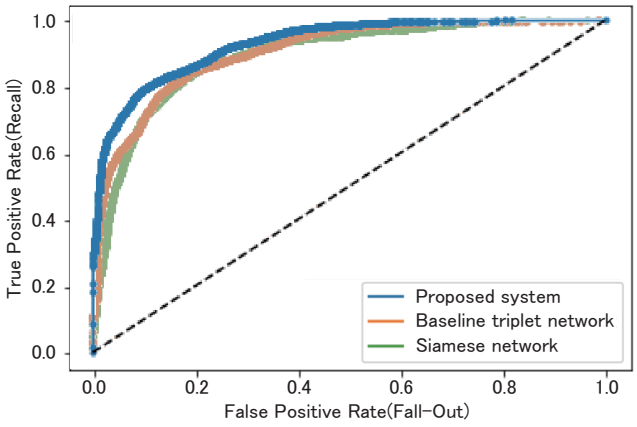
Table 3. Performance benchmarking with respect to the best EER (%) averaged from five runs of two-fold cross-validation test on Wi-Fi CSI signature dataset.

Methology	Best EER (%)	Condition
LSE	48.44	-
PCA-LSE	30.79	Reduced dimension=40
SVM (Linear)	28.23	c=1
SVM (RBF)	24.31	c=1, γ=0.01/3000
TER-RM2	35.84	M=1,τ=η=0.5
Siamese network	23.53	lr=0.00005
Baseline triplet network	20.34	lr=0.00005, α=0.1
Proposed system	**19.35**	**lr=0.00005, α=0.1**

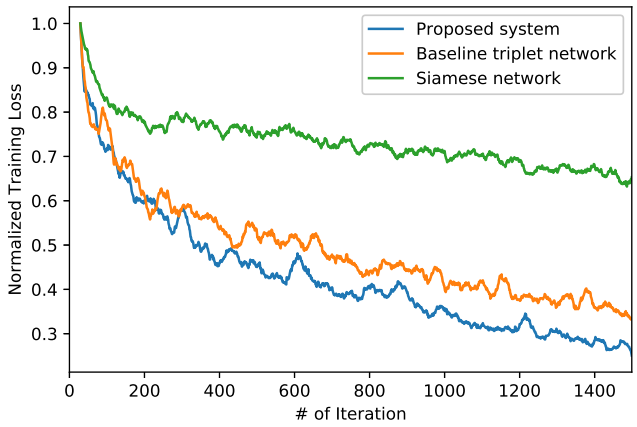
As shown in the Fig. 2(b), the proposed system clearly shows the largest Area Under Curve (AUC) among three compared methods while other two methods show similar AUC. Moreover in Fig.2(b) which illustrates the training loss curve along the number of training iteration, the proposed system shows the fastest training loss convergence followed by the baseline triplet network. We note here that 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 based on their function. According to these two observations, it can be concluded that the triplet input mining with KAR space learning improves not only the training loss convergence speed but also the verification performance.

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). The kernel and the range space learning was adopted for refining the triplet inputs for fast loss convergence by mining the distinctive inputs from the training Wi-Fi signature signals. Subsequently, the triplet network utilizing the ConvNet structure was trained with the mined triplet inputs based on



(a) ROC Curve



(b) Normalized training loss curve

Fig. 2. (a) shows the Receiver Operating Characteristic(ROC) Curve and (b) shows the normalized training loss curve of the deep learning based methods.

L-2 distance comparison. Our experiments on an in-house Wi-Fi handwritten signature dataset showed encouraging verification accuracy with faster training loss convergence compared with the baseline triplet network and the Siamese network.

References

1. Bailador, G., Sanchez-Avila, C., Guerra-Casanova, J., de Santos Sierra, A.: Analysis of pattern recognition techniques for in-air signature biometrics. Pattern Recognition **44**(10-11), 2468–2478 (2011)
2. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: a deep quadruplet network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 403–412 (2017)
3. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: 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. pp. 1335–1344 (2016)
4. Ding, S., Lin, L., Wang, G., Chao, H.: Deep feature learning with relative distance comparison for person re-identification. Pattern Recognition **48**(10), 2993–3003 (2015)
5. Galbally, J., Diaz-Cabrera, M., Ferrer, M.A., Gomez-Barrero, M., Morales, A., Fierrez, J.: On-line signature recognition through the combination of real dynamic data and synthetically generated static data. Pattern Recognition **48**(9), 2921–2934 (2015)
6. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)
7. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)
8. Hoffer, E., Ailon, N.: Deep metric learning using triplet network. In: International Workshop on Similarity-Based Pattern Recognition. pp. 84–92. Springer (2015)
9. Jeon, J.H., Oh, B.S., Toh, K.A.: A system for hand gesture based signature recognition. In: 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV). pp. 171–175. IEEE (2012)
10. Koch, G., Zemel, R., Salakhutdinov, R.: Siamese neural networks for one-shot image recognition. In: ICML deep learning workshop. vol. 2 (2015)
11. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
12. Malik, J., Elhayek, A., Ahmed, S., Shafait, F., Malik, M., Stricker, D.: 3DAirSig: A framework for enabling in-air signatures using a multi-modal depth sensor. Sensors **18**(11), 3872 (2018)
13. Moon, H.C., Jang, S.I., Oh, K., Toh, K.A.: An in-air signature verification system using Wi-Fi signals. In: Proceedings of the 2017 4th International Conference on Biomedical and Bioinformatics Engineering. pp. 133–138. ACM (2017)
14. Pokkunuru, A., Jakkala, K., Bhuyan, A., Wang, P., Sun, Z.: Neuralwave: Gait-based user identification through commodity WiFi and deep learning. In: IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society. pp. 758–765. IEEE (2018)
15. Sanmorino, A., Yazid, S.: A survey for handwritten signature verification. In: 2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering. pp. 54–57. IEEE (2012)
16. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)
17. Sesa-Nogueras, E., Faundez-Zanuy, M., Mekyska, J.: An information analysis of in-air and on-surface trajectories in online handwriting. Cognitive Computation **4**(2), 195–205 (2012)

18. Shi, C., Liu, J., Liu, H., Chen, Y.: 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. p. 5. ACM (2017)
19. Toh, K.A.: Fingerprint and speaker verification decisions fusion. In: 12th International Conference on Image Analysis and Processing, 2003. Proceedings. pp. 626–631. IEEE (2003)
20. Toh, K.A.: Kernel and range approach to analytic network learning. International Journal of Networked and Distributed Computing **7**(1), 20–28 (December 2018). https://doi.org/https://doi.org/10.2991/ijndc.1970.1.7.3
21. Toh, K.A.: Learning from the kernel and the range space. In: 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS). pp. 1–6. IEEE (2018)
22. Toh, K.: Analytic network learning. arXiv preprint arXiv:1811.08227 (November, 2018)
23. Toh, K.A., Eng, H.L.: Between classification-error approximation and weighted least-squares learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(4), 658–669 (2008)
24. Toh, K.A., Lin, Z., Li, Z., Oh, B., Sun, L.: Gradient-free learning based on the kernel and the range space. arXiv preprint arXiv:1810.11581 (2018)
25. Wang, F., Zuo, W., Lin, L., Zhang, D., Zhang, L.: 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. pp. 1288–1296 (2016)
26. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In: Advances in neural information processing systems. pp. 1473–1480 (2006)