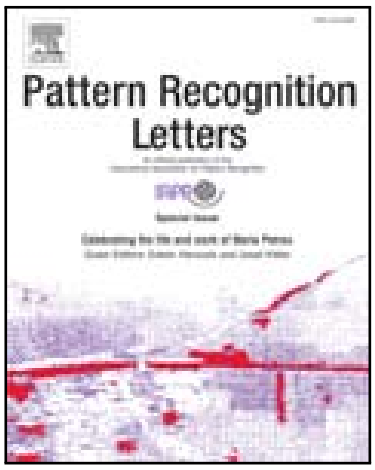

Are mid-air dynamic gestures applicable to user identification?

Heng Liu , Liangliang Dai , Shudong Hou , Jungong Han , Hongshen Liu

PII: S0167-8655(18)30146-6
DOI: 10.1016/j.patrec.2018.04.026
Reference: PATREC 7154

To appear in: *Pattern Recognition Letters*

Received date: 1 February 2018
Revised date: 15 April 2018
Accepted date: 17 April 2018

Please cite this article as: Heng Liu , Liangliang Dai , Shudong Hou , Jungong Han , Hongshen Liu , Are mid-air dynamic gestures applicable to user identification?, *Pattern Recognition Letters* (2018), doi: 10.1016/j.patrec.2018.04.026

Highlights

- We investigate the feasibility of mid-air dynamic gesture based user identification by providing a Bi-GRU network.
- We reveal the relationship between the gesture type features and the gesture user identity characteristics.
- We explore the availability of ITQ hash coding on gesture based user identification and gesture classification.
- We discuss the effect of different gestures on the performance of user identification through hash ITQ feature representation.

Are mid-air dynamic gestures applicable to user identification?

Heng Liu*[1, 2], Liangliang Dai[1], Shudong Hou[1], Jungong Han[3], Hongshen Liu[1]

[1] *Anhui University of Technology, Maxiang Road, Ma'anshan 243032, China*
[2] *Key Laboratory of Intelligent Perception and Systems for High Dimensional Information of Ministry of Education, Nanjing 210094, China*
[3] *Lancaster University, LA1 4YW, U. K.*

ABSTRACT
Unlike the existing gesture related research predominantly focusing on gesture recognition (classification), this work explores the feasibility and the potential of mid-air dynamic gesture based user identification through presenting an efficient bidirectional GRU (Gated Recurrent Unit) network. From the perspective of the feature analysis from the Bi-GRU network used for different recognition tasks, we make a detailed investigation on the correlation and the difference between the gesture type features and the gesture user identity characteristics. During this process, two unsupervised feature representation methods – PCA and hash ITQ (Iterative Quantization) are fully used to perform feature reduction and feature binary coding. Experiments and analysis based on our dynamic gesture data set (60 individuals) exemplify the effectiveness of the proposed mid-air dynamic gesture based user identification approach and clearly reveal the relationship between the gesture type features and the gesture user identity characteristics.
Keywords: Gesture based user identification, gesture user identity characteristics, Bi-GRU, mid-air dynamic gestures

* Corresponding author. Tel.: +0-86-555-2315543; fax: +0-86-555-2315543; e-mail: hengliusky@aliyun.com

1. Introduction

User biometrics refers to the automatic recognition of user identity based on physiological or behavioral characteristics (e.g., face recognition [1], fingerprint recognition [2], iris recognition [3], and handwriting recognition [4]). Since physiological characteristics based biometric technology is unique portability and not easily lost, it becomes a widely acknowledged user identification means. While relative to the fingerprints, iris and other static physiological characteristics, the individual's behavioral characteristics are dynamic, variable, more difficult to imitate and counterfeit. Therefore, user identification based on dynamic biological characteristics draws more and more research attention in recent years. such as gait recognition [4], full body motion identification [5], biometric authentication based on mouse movements and typing rhythm [6, 7].

Recently, mid-air dynamic gesture based biometrics is proposed in works [8, 9, 10]. The basic principle of this biometrics means is that even different individuals draw the same type gestures, there will be inherent motion differences due to behavioral habits. According to such kind of unique diversity information, the identities of different individuals can be determined. In particular compared to other user biometrics, an attractive advantage of gesture based user identification is that the gesture itself can certainly express sufficient motion or action information which can be recognized at the same time. In addition, gesture recognition (classification research has made great progress [11, 12, 13, 23, 24, 25] in recent years. Thus dynamic gesture based user biometrics not only can integrate human-computer gesture interaction but also can effectively realize user identification.

Most current gesture based user biometrics works [9, 10] mainly focus on specific tasks and are under small samples data sets and they still employ the handcrafted features and adopt the traditional methods – DTW (dynamic time warping) or SVM (support vector machine) for user identity matching. Furthermore, these kind of methods distinguish individuals only by comparing the dynamic gesture trajectories without considering the correlation and the relationship between the gesture motion features and the gesture characteristics for user identity. When more individuals and more gestures are involved, the motion trajectories of user gestures will be partially overlapped and personalized, then the existing dynamic gesture based user identification ways will be challenged.

In view of this, in this work, we explore mid-air dynamic gesture based user biometrics by analyzing the relationship between the gesture classification and the gesture based user identification. We firstly utilize Microsoft Kinect to capture 3D gesture motion information and extract 25 body joints as the mid-air dynamic gesture sequences data. Then based on these gesture sequences, for different recognition tasks (gesture classification and gesture based user identification), we present to expand the GRU (Gated Recurrent Unit) network [15, 16] to form bidirectional GRU (Bi-GRU) model to learn the gesture category features and the gesture user identity characteristics, respectively. Our contribution can be summarized as follows:

- By providing an efficient Bi-GRU network based user identification framework, we make a detail investigation on the feasibility of mid-air dynamic gesture based user identification.
- For the first time, we reveal the correlation and the difference between the gesture type features and the gesture user identity characteristics. We also provide the detail performance assessments and the comparisons with other state-of-the-art methods.
- We explore the effectiveness of hash ITQ coding for gesture based user identification and gesture classification. We also discuss the effect of different types gestures on the performance of user identification.

The rest of the paper is organized as follows. Section 2 describes the procedure of gesture data acquisition and preprocessing in detail. Section 3 provides an improved DTW method and the Bi-GRU network based approach for user identification. The feature extraction and representation techniques are also introduced in this section. In section 4, lots of experimental results and the analysis of the feasibility of the proposed gesture user identification are provide. And the correlation and the difference between gesture category features and the gesture user identity characteristics are also focused in this section. Finally, we conclude this work in Section 5.

2. Mid-air dynamic gestures acquisition and preprocessing

We utilize Microsoft Kinect to capture mid-air dynamic gestures and record hands motion trajectories. In order to reveal the relationship between gesture categories and the user identities, in this work, three kinds of gestures were preset and required to be performed by all subjects. These pre-set gestures include right hand mid-air drawing 'O', left hand drawing 'V' and two hands clapping (they will be shortly noted as 'O', 'V' and clapping). The Kinect can provide skeleton node data which contains 25 main-body joints, including spine-base, spine-mid, neck, head, spine- shoulder, as well as the left- and right-side joints - shoulder, elbow, wrist, hand, hip, knee, ankle, foot, hand-tip and thumb. Actually, each gesture data consists of all coordinates (x, y and z) of skeletal joints during gesture motion.

The original gesture sequences should be preprocessed to remove the interference of body translational motion and joint jitter. Once the gesture sequence is smoothed and normalized, gesture spotting should be performed to determine which frame correspond to the motion start and which to the motion end. By this way, all successive dynamic gestures can be dissembled into the short and independent motion sequences. Then the gesture sequences will be piped into the proposed Bi-GRU network for user identification.

2.1. Data preprocessing

There involve two operations in data preprocessing: normalization and motion noise removal. The aim of data normalization is to overcome the natural data biases from capture device. In the scenario, most biases arise from body translation motion and the physical size changing of certain subjects. To rectify this, spine-based length normalization was applied as follows:

$$p^g_{center,i,t} = \left(X^g_{i,t} - X^g_{spine,t}, Y^g_{i,t} - Y^g_{spine,t}, Z^g_{i,t} - Z^g_{spine,t}\right) \quad (1)$$

$$p^g_{norm,i,t} = \frac{p^g_{center,i,t}}{\left\| p^g_{center,neck,t} - p^g_{center,spine,t} \right\|} \quad (2)$$

where $X^g_{i,t}, Y^g_{i,t}, Z^g_{i,t}$ are the 3D coordinates of joint node p_i at time t for gesture g. We firstly take the skeleton spine node as the root node to center a gesture, which the relative positions between all joint nodes and the root node are adopted at every time. In order to normalize the size of the observed subject, all coordinates are scaled according to the distance between the neck and the spine joints.

Once the normalized joint positions are obtained, we take Gaussian smoothing along the temporal dimension to reduce the skeleton motion noise. Suppose there are five adjacent points such as $p_{i-2}, p_{i-1}, p_i, p_{i+1}, p_{i+2}$, we can determine the value of the standard deviation of Gauss distribution as:

$$\delta = max\begin{pmatrix}\sum\|p_{i-2} - p_{i-1}\| + \|p_{i-1} - p_i\| \\ \sum\|p_i - p_{i+1}\| + \|p_{i+1} - p_{i+2}\|\end{pmatrix} \quad (3)$$

Then we can get new smoothed joint nodes as:

$$p_i^{new} = \frac{\sum_{j=i-2}^{i+2} G(d_j)p_j}{\sum_{j=i-2}^{i+2} G(d_j)} \quad (4)$$

where $d_j = \sum_k \|p_k - p_{k-1}\|$ and $G(\cdot)$ is the Gaussian function.

2.2. Gesture spotting

The normalized and noise free gestures will be continually carried out gesture spotting to extract the genuine gesture sequences. Aiming for real time gesture based biometrics, the basic requirement for gesture spotting is that it should be fast and robust, which will be beneficial for gesture based user identification or gesture recognition. In this work, we take an ELM (extreme learning machine) classification technique for fast and robust gesture sequences location. More related information about ELM can refer to the work [14]. With such gesture detection technique, the gesture sequences can be accurately determined.

It should be noted that all gesture sequences will be scaled to the same sequence length N in practice. Given one gesture sequence which contains L frames, the scaling method can be described as:

$$index_i = \frac{L}{N} \times i \quad (5)$$

where $index_i$ is the index of the ith sampled frame. Then, a normalized and scaled gesture sequence can be represented as:

$$G = (index_1, index_2, \dots, index_N) \quad (6)$$

3. The proposed gesture based user identification approach

3.1. An improved DTW method

DTW algorithm is a nonlinear structured technique which combines the distance measuring with time warping. Since DTW can establish a reasonable alignment path between the test signal and the reference pattern, it is widely used in sequence signal matching, especially for action recognition and speech recognition. DTW algorithm will try to find an optimal path which can make the path cost function (usually is Euclidian distance) to get the minimum value. DTW gesture sequences matching can be formulated as:

$$\mathrm{DTW}(G_{ti}^{g1}, G_{tj}^{g2}) = \min(\sum_p \|G_{p,ti}^{g1} - G_{p,tj}^{g2}\|) \quad (7)$$

where G_{ti}^{g1}, G_{tj}^{g2} are two different gesture sequences which have been preprocessed and p is the joint node index.

To enhance the matching performance, we take gesture template synthesis technique – super gesture template [8] when applying DTW for gesture based user identification or gesture categorization. In practice, we take two gesture templates to form super template. Assuming the two gesture sequences are $\mathrm{X} = (x_1, x_2, \dots, x_i, \dots)$ and $\mathrm{Y} = (y_1, y_2, \dots, y_j, \dots)$ respectively, through DTW matching between these two sequences we can get the warp path $\mathrm{W} = (w_1, w_2, \dots, w_r, \dots)$, then the super gesture template SG can be acquired by:

$$SG = (sg_1, sg_2, \dots, sg_r, \dots) \quad (8)$$

where $sg_r = \frac{(x_i + y_j)}{2}$.

3.2. Bi-GRU network

GRU [15, 16] network is an effective technique for sequence signal learning, classification and prediction. The basic architecture of GRU is the same as LSTM except that the hidden layer updates are replaced by purposed built memory cells. The architecture of a GRU unit is shown in Figure 1.

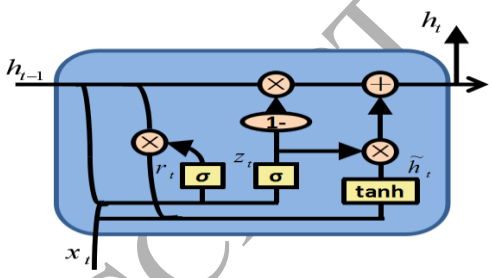

Fig.1. The architecture of a GRU unit

According to the architecture, an GRU unit accepts an input sequence $\mathrm{x} = (x_1, x_2, \dots, x_t)$, and calculates an output sequence $\mathrm{y} = (y_1, y_2, \dots y_t)$. GRU can produce the current time cell control state h_t from the update gate z_t and the reset gate r_t. The update gate decides how much the unit updates its activation or content. And the reset gate effectively makes the unit act as if it is reading the first symbol of an input sequence, allowing it to forget the previously computed state. The detail implementation of a GRU cell can be formulated as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (9)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (12)$$

Where σ is the logistic sigmoid function, r_t, z_t and h_t are the reset gate, update gate and status information.

One-way GRU can be extended to form bidirectional GRU (Bi-GRU) by taking backward time sequence into consideration. The most prominent advantage of Bi-GRU is that it not only can learn the forward information like traditional GRU but also can introduce the backward information. This actually means that the output of the input at time t depends on the both outputs at $t-1$ and $t+1$. This kind of mechanism is more conducive to the sequence signal based feature learning. Our Bi-GRU network based user identification model is shown in Figure 2.

In the model, the preprocessed gesture skeleton data is treated as the inputs of the following neural network layers. The first full connection layer is used to extract the spatial features, and the forward and backward temporal dependence information are learned by a Bi-GRU layer. Then spatiotemporal gesture features can be obtained by integrating the bidirectional sequence information. Finally, gesture categories or user identities may be recognized by taking the second full connection layer and after soft-max classification.

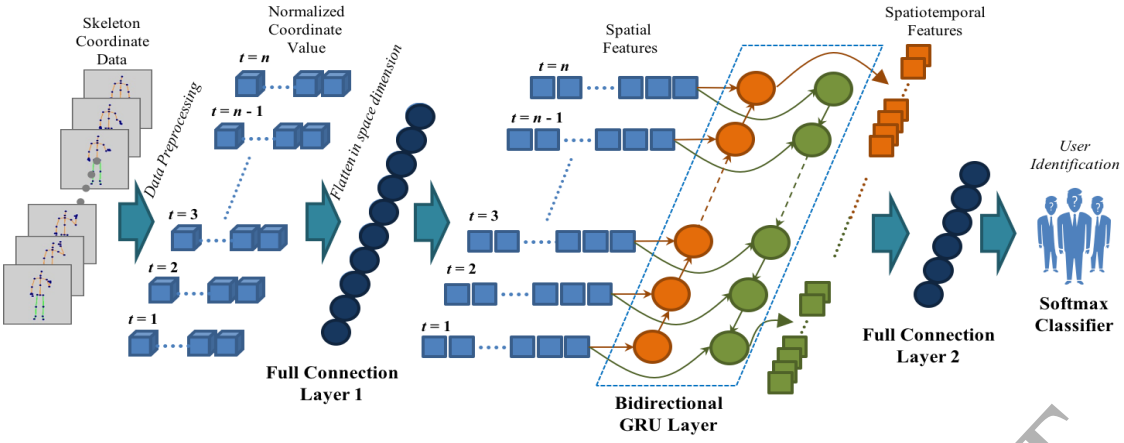

Fig. 2. The proposed Bi-GRU network gesture based user biometrics model

3.3. Gesture based user identification vs. gesture classification

Input with diverse categories of data, the proposed Bi-GRU network can be trained to learn different kinds of sequences spatiotemporal features. In order to analyze the correlation between gesture type features and gesture user identity characteristics, gesture labels and identity labels are used to train Bi-GRU networks for different tasks, respectively. For a fair comparison, the network structures and initial parameters are set to be the same in such two different tasks.

In the proposed Bi-GRU recognition model, the output number in the first full connection layer is 512, and the number of processing GRU units in the forward layer and the backward layer are both 512. Therefore, after concatenating the dimension of spatiotemporal gesture features is 1024. The output number of second full connection layer for user identification task is 60 (60 individuals) whereas for gesture categorization task it becomes to be 3 (3 gesture categories). Thus, through the Bi-GRU network, one gesture sequence can be transformed into two 1024-dimensional features: gesture category feature and gesture user identity characteristic.

Feature compact representation may facilitate parsing the essential nature and the effective pattern matching. In this work, we use the classical PCA and a hash binary coding method – ITQ (iterative quantization) [17] to achieve the features compact representation after they are extracted from Bi-GRU network. Actually, ITQ is utilized to find an optimal rotation matrix R in multi-dimensional space to minimize the binary quantization error Q between the dimension reduction gesture feature V and the Hypercube B (the binary code):

$$Q(B,R) = \min\|B - VR\|^2 \tag{13}$$

Through feature compact representation, the gesture category features and the gesture user identity characteristics can be easily analyzed and visualized. In addition, especially for ITQ binary presentation, gesture categories or user identities may be quickly recognized via Hamming distance matching.

In addition, since the intra-class variance can be used to describe the dispersion in the same feature pattern and the inter-class difference can describe the degree of difference between different categories feature patterns, we use these two criteria to observe and evaluate the correlation and the difference between the gesture features for classification and the ones for user biometrics. Assuming the two different categories of gesture feature patterns are represented as X_i^{c1} $(i = 1, 2, \dots, n)$ and X_j^{c2} $(j = 1,2, \dots, m)$, the superscripts indicate the category and the subscripts indicate the index of each sample. Then the intra-class variance Var can be expressed as:

$$Var = \frac{\sum_{i=1}^{n}(\|X_i - \bar{X}\|)^2}{n} \tag{14}$$

where $\bar{X}$ represents the mean of the category. The inter-class distance Dis of two categories can be described as:

$$Dis = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\left\|X_i^{c1} - X_j^{c2}\right\|\right)^2}{n \cdot m} \tag{15}$$

4. Experiments and analyses

4.1. Dataset and training details

Since there is no public and available mid-air dynamic gesture based biometrics data set according to the best of our knowledge, we use Microsoft Kinect to capture three types mid-air dynamic gestures of 60 individuals for user identification. Among these individuals, there are 21 females and 39 males with body height varying from 1.5 meters to 1.9 meters, and body weight changing from 45 Kg to 90 Kg. Three types of mid-air dynamic gestures that can be performed by different hands were pre-set, which include right hand drawing 'O', left hand drawing 'V' and two hands clapping. Each individual was captured these three types gestures 20 times. Then, our mid-air dynamic gesture biometrics dataset contains totally 3600 mid-air dynamic gestures from 60 individuals. [1]

The proposed Bi-GRU network is trained based on above gesture data set with Google Tensorflow platform. And the weights and biases of the network are all randomly initialized. Before training, the learning rate is set to be 0.0005 and ADAM optimization is used as a solver with a batch size of 256. Two NVIDIA 1080ti GPUs are consumed for the training and it needs about 2500 iterations before convergence, which totally takes about 20 minutes.

4.2. Performance on gesture based user identification

We randomly take 14 gesture samples as training data and the last 6 as test data and all the gesture sequences were scaled to 65 frames. In each frame, the 3D coordinates of all the 25 body joints are taken as gesture data. This means a single gesture can be described by 4875 $(65 \times 25 \times 3)$ coordinate values. The scenario of user identification includes two situations: 1) each of three types gestures is used for user identification; 2) three types of gestures are mixed as one kind of gesture input for user

[1] https://drive.google.com/open?id=1G_vxXDDn37VB3CFj3q1fmTK0imfq7pC6

identification. In Table. 1, we compare our Bi-GRU network based user identification approach not only with traditional methods - SVM, DTW but also with the state of the art deep learning methods – LSTM, GRU, etc.

As shown in Table 1, the proposed Bi-GRU network achieves the best performance. The user identification performances of three types of gestures are all very high – drawing 'O', drawing 'V and 'Clapping' will get 99.44%, 99.44% and 98.89%, respectively. In addition, using the mixed type gestures for user identification can also reach 99.35%. All these reveals that gesture based user identification is feasible and do not depend on any special type gesture in terms of such three kinds of gestures.

Table 1. Performance on dynamic gesture based user identification

Gestures / Methods	'O'	'V'	Clapping	Mixed Gestures	Average
SVM	0.9111	0.9250	0.9361	0.8472	0.9048
DTW	0.9894	0.9863	0.9742	0.9833	0.9833
LSTM	0.9833	0.9778	0.9694	0.9842	0.9786
GRU	0.99117	0.9889	0.9806	0.9851	0.9865
Bi-LSTM	0.9917	0.9917	0.9833	0.9898	0.9891
Bi-GRU	**0.9944**	**0.9944**	**0.9889**	**0.9935**	**0.9928**

In order to investigate whether the performance of dynamic gesture based user identification is also affected by users' body motion, we only choose 10 hand joints (elbow, wrist, hand, hand-tip and thumb) and re-implement the user identification experiments. The results are shown in Table 2.

Table 2. Performance on only using hand joints

Gestures / Methods	'O'	'V'	Clapping	Mixed Gestures	Average
SVM	0.8444	0.8889	0.8583	0.7630	0.8386
DTW	0.5864	0.5833	0.4152	0.5283	0.5283
LSTM	0.9556	0.9333	0.9028	0.9462	0.9335
GRU	0.9667	0.9694	0.9444	0.9593	0.9599
Bi-LSTM	0.9611	0.9528	0.9333	0.9509	0.9495
Bi-GRU	**0.9861**	**0.9750**	**0.9500**	**0.9675**	**0.9696**

From Table 2, it is obvious without body motion information, the performance of the proposed way will drop slightly. Actually, From Table 1 and Table 2, we can easily get two points: mid-air dynamic gestures are really applicable for user identification; the dynamic gestures used for user identification involve the motion of all body joints including hand joints. However, for accurate discussion and fair play, in all the following experiments, we do not use the motion information of whole body joints and only take the hand joints' motion sequences as dynamic gestures data.

4.3. Gesture user identity characteristics vs. Gesture category features

A) Can gesture user identity characteristics be used for gesture classification?

Here, we manage to investigate whether gesture user identity characteristics can also be applied for gesture classification. That is, we manage to examine whether the gesture identity characteristics also contain the gesture types information. In order to ensure the credibility, the same training data (see the first paragraph of Section 4.2) is used to train the proposed Bi-GRU network and supervised by gesture categories and user identities, respectively. Thus, after training, for train data we get 2520 $(14 \times 3 \times 60)$ gesture category features and 2520 gesture based user identity features. As for test data, we also can get 1080 $(6 \times 3 \times 60)$ gesture category features and 1080 user identity characteristics. It is worth noting that each feature vector takes 1024 dimensions.

In order to facilitate the analysis, we take PCA technique for features dimension reduction. Based on gesture category clustering, Figure 3 shows the distributions of user identity characteristics (left) and the gesture type features (right), respectively in the dimensionality reduction space (3D space). The intra-class variance and the inter-class difference of such two kinds of gesture features are given in Table 4.

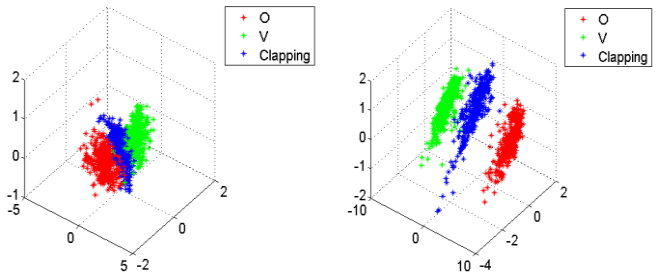

Fig. 3. The distribution of gesture based user identity characteristics (left) and the gesture type features (right) based on gestures clustering.

Table 4. Intra-class variances and inter-class difference of gesture type features and gesture user identity characteristics under gesture classification

Features / Gestures	Intra-class variances			Inter-class difference		
	'O'	'V'	Clapping	'O' – 'V'	'V' - Clapping	'O' - Clapping
User identity characteristics	0.6221	**0.5700**	0.7763	**2.8212**	1.7568	1.8064
Gesture type features	**0.8375**	0.9131	1.3894	**134.7118**	61.3632	20.2139

From the Figure 3 and Table 4, it is clear that compared to gesture type features, although the intra-class variance and the inter-class difference of gesture user identity characteristics are both smaller, the distribution of its gesture pattern is not overlapped. This indicated that the gesture based user identity characteristics also contains certain degree or 'weak' gesture category information, even it is more suitable for user biometrics.

In order to re-verify whether the gesture user identity characteristics can play a role in gesture classification, we manage to perform K-NN classification. PCA is still used to reduce features' dimensionality. For each user identity characteristic in test data, the K nearest samples of training features in Euclidean space were selected to vote the gesture category of the test sample. Also each gesture category features in test set is handled in the same way. The average gesture recognition results are shown in Table 5.

Table 5. K-NN based gesture classification

K / Dim	Gesture user identity characteristics			Gesture type features		
	K=1	K=3	K=5	K=1	K=3	K=5
1	0.3426	0.2769	0.2556	0.9907	0.9954	0.9954
3	0.8731	0.8963	0.9000	0.9991	0.9981	0.9981
5	0.9417	0.9463	**0.9491**	0.9991	0.9991	**0.9981**

Through the experimental results in the table, it can be found that for the task of gesture user identity characteristics still can achieve good performance – 94.91% if configured with the simple 5-NN classifier. All experimental results in Table 4 and Table 5 demonstrate that gesture user identity characteristics not only can be used for user identification, but also can be applied for gesture classification.

B) Can gesture type features be used for user identification?

In this part, we mainly discuss whether the gesture type features contain the unique users' identities information. For the convenience of illustration and demonstration, we randomly

selected 3 individuals and choose the right hand gesture 'O' for comparison and discussion. Figure 4 shows the distributions of gesture type features (left) and gesture user identity characteristics (right) based on three individuals' identities clustering. Also, under the task of user identification, the intra-class variance of each user identity and the inter-class difference between two users' identities are given in Table 6.

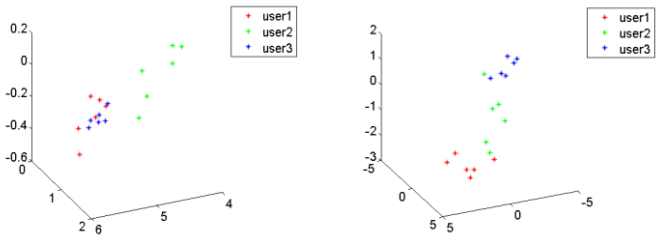

Fig. 4. The distribution of gesture type features (left) and the user identity characteristics (right) based on users' identities clustering.

Table 6. Intra-class variances and inter-class difference of gesture type features and gesture user identity characteristics under user identification

Features \ Users	Intra-class variances			Inter-class difference		
	User1	User2	User3	User1 - User2	User2 - User3	User1 - User3
Gesture type features	0.0675	0.1375	**0.0211**	**1.1921**	0.9762	0.0979
Gesture user identity characteristics	1.2468	2.0385	**0.6190**	24.8104	7.1171	**38.7541**

From the Figure 4, we can see that the gesture characteristic patterns of the same identify is really grouped together while the gesture type features are cluttered. Then according to Table 6 the inter-class difference of gesture type features is very small. This means that gesture type features do mix the identities of multiple individuals.

For the sake of re-investigate the feasibility of taking gesture type features for user identification, K-NN classifier is used to performed classification again. For the same gesture 'O', there are totally 1200 (60×20) gesture type features, of which 360 features were taken as test data and the remaining 840 features were used as training data. For fair play, user identify chars are handled in the same way. The identity of each test feature is determined by the voting of the K-nearest neighbors in Euclidean space. Finally, the average user identification results are shown in Table 7.

Table 7. K-NN based user identification

Dim \ K	Gesture type features			Gesture user identity characteristics		
	K=1	K=3	K=5	K=1	K=3	K=5
1	0.0528	0.0194.	0.0139.	0.0583	0.0306	0.0139
3	0.2972	0.2417	0.2194	0.5194	0.4611	0.4278
5	0.5389	0.4500	0.4194	0.7583	0.7556	0.7083
50	**0.8528**	0.7833	0.7250	**0.9583**	0.9472	0.9278

From Table 7, we can see that the user identification performance by using gesture category features is quite limited when its dimensionality is not high. In addition, if K (the number of the nearest neighbors) increases, the user identification performance drops much. This re-verify the meanings of Figure 4 (left) that the users' identities in the gesture category features are badly mixed in low dimension space. In addition, for gesture based user identity characteristics, only when the feature dimensionality reaches up to 50, the identification performance can achieve more than 95%. The fact means that the distribution of gesture user identity characteristics is complicated and distributed in high dimensional space. All the experimental results and the comparisons confirm that the gesture type features are not conducive to direct user identification due to its mixture of different individuals' identities.

4.4. Gesture based user identification with hash coding

For large scale mid-air dynamic gesture data set, how to quickly match and query some specific test gestures for user identification is a worth studying problem. Fortunately, hash binary coding methods, such as ITQ [17] and recent deep hash models [18-22], throw much light on this challenge. In this part, we will research the effectiveness of hash feature representation for gesture based user identification and gesture classification. Specifically, in this work, the convenient hash coding method – ITQ [17] will be adopted to convert gesture features or user identity characteristics into binary codes. Then Hamming distance is utilized for feature patterns matching.

The experimental configuration is the same as before. 14 gestures of each type gesture are selected as training samples and the remaining 6 samples are treated as test data. For gesture classification, finally we get 2520 training samples and 1080 test samples. Then they are firstly converted into a 1024-dimensional gesture type feature vectors by the proposed Bi-GRU network. After normalization, these feature vectors are continually converted into 16-bit binary numbers by ITQ coding. Table 8 shows the performance of gesture classification based on ITQ coding.

Table 8. ITQ coding based gesture classification

Gesture labels	'O'	'V'	'Clapping'
Recognition accuracy	1.000	1.0000	0.9600

For user identification, here we can investigate whether the different categories gestures hold different recognition performance. Similarly, there are totally 840 (60×14) user identity training samples and 360 test samples for each type of gesture. And they are also converted into 1024-dimensional feature vectors. It has been shown in Section 4.4.3 (B) that gesture based user identity characteristics is complicated and need more dimensions to be expressed, thus here the user identity characteristics are represented with 64 bit binary codes by ITQ coding. Table 9 shows the user identification performance based on ITQ coding for different types gestures. Figure 5 shows the distribution of 24-bits ITQ hash coding for seven user identities characteristics under gesture type 'O' (every 8-bits translated into one decimal number and treated as a dimension coordinate value).

Table 9. User identification of each type gesture with ITQ coding

Gesture types	'O'	'V'	'Clapping'
Recognition accuracy	0.8139	0.8083	0.7083

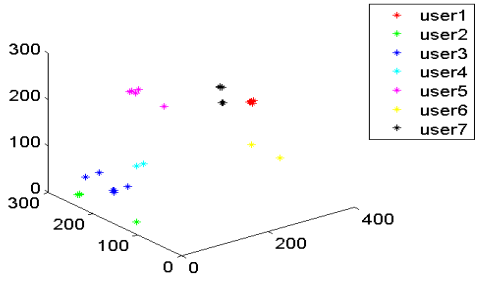

Fig. 5. The distribution of seven user identities characteristics based on ITQ hash coding

From Table 8, it is obvious that the same type gesture can be mapped into a very similar even unique hash code by ITQ representation method. This means with Hamming distance matching, the gesture types can be quickly and effectively distinguished. On the other hand, from Table 9 and Figure 5, we can get that two points: right hand gesture 'O' holds the strongest potential to discriminate user identities; after hash coding different gesture identities characteristics are basically represented into different identities clusters and all the user identification performance based on this are not bad. All these experiments indicate that both the gesture user identity characteristics and the gesture type features can be effectively represented by hash coding to achieve fast and good recognition results.

5. Conclusion

In this work, we make a detail investigation and declare that the mid-air dynamic gestures are applicable to user biometrics. We collected a mid-air dynamic gesture data set from 60 individuals with three types of gestures. We propose an efficient Bi-GRU network based model to perform gesture based use identification and will achieve about 97% rank-one recognition accuracy even if only hand joints' motion is considered as dynamic gesture.

Furthermore, we manage to explore the correlation and difference between gesture classification and gesture based user identification. By using the proposed Bi-GRU network to extract gesture type features or gesture identity characteristics, we also make a focus discussion on whether the gesture type features and gesture based user identity characteristics can be applied to each other's recognition tasks. We found that the gesture based user identity characteristics are complicated detail features and contain the available gesture types information, then it can be used for gesture classification. Meanwhile, the gesture type features mix the identities information of multiple individuals and are not suitable for user identification directly.

Finally, we find that ITQ hash coding is effective for the representations of gesture user identity characteristics and the gesture type features. In addition, we get to know that different types gestures hold different matching performance on gesture based user identification.

Acknowledgement

This work is supported by the Major Project of Natural Science of Anhui Provincial Department of Education (Grant No. KJ2015ZD09) and by the Anhui Provincial Natural Science Foundation (Grant No. 1608085MF129). It is also supported by the Innovation Foundation of Key Laboratory of Intelligent Perception and Systems for High Dimensional Information of Ministry of Education (Grant No. JYB201705).

References

1. Abate A F., et al., 2007. 2D and 3D Face Recognition: A Survey. Pattern Recognition Letters, 28(14), 1885-1906.
2. Maltoni D., et al., 2009. Handbook of Fingerprint Recognition. Springer Science & Business Media.
3. Nabti M., Bouridane A., 2008. An Effective and Fast Iris Recognition System based on a Combined Multiscale Feature Extraction Technique. Pattern Recognition, 41(3), 868-879.
4. Sebastijan S., Matjaz B J., 2015. Inertial Sensor-Based Gait Recognition: A Review. J. Sensors (Basel, Switzerland), 15(9), 22089-22127.
5. Munsell B C., et al., 2012. Person Identification Using Full-body Motion and Anthropometric Biometrics from Kinect Videos. European Conference on Computer Vision. 91-100.
6. Ahmed A A E., Traore I., 2012. A New Biometric Technology based on Mouse Dynamics. IEEE Transactions on Dependable and Secure Computing, 4(3), 165.
7. Banerjee S P., Woodard D L., 2012. Biometric Authentication and Identification Using Keystroke Dynamics. Journal of Pattern Recognition Research, 7(1): 116-139.
8. Guna J., et al., 2014. User Identification Approach based on Simple Gestures. Multimedia Tools and Applications, 71(1), 179-194.
9. Zhang H., et al., 2014. A Personalized Gesture Interaction System with User Identification Using Kinect. Pacific Rim International Conference on Artificial Intelligence. Springer, Cham, 614-626.
10. Mendels O., et al, 2014. User Identification for Home Entertainment based on Free-air Hand Motion Signatures. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44(11), 1461-1473.
11. Neverova N, et al, 2014. A Multi-scale Approach to Gesture Detection and Recognition. IEEE International Conference on Computer Vision Workshops. IEEE, 484-491.
12. Neverova N, et al, 2016. ModDrop: Adaptive Multi-Modal Gesture Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 38(8):1692-1706.
13. Rautaray S S., Agrawal A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. Kluwer Academic Publishers.
14. Liu H., et al., 2017. Extreme Learning Machine and Moving Least Square Regression Based Solar Panel Vision Inspection. Journal of Electrical and Computer Engineering.
15. Dey R, Salemt F M, 2017. Gate-variants of Gated Recurrent Unit (GRU) neural networks. IEEE International Midwest Symposium on Circuits and Systems. IEEE, 1597-1600.
16. Chung J, et al., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv: 1412.3555.
17. Y Gong, et al., 2011. Iterative Quantization: A Procrustean Approach to Learning Binary Codes. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).
18. F. Shen, Y. Xu, L. Liu, Y. Yang, Z. Huang and HT. Shen, 2018. Unsupervised Deep Hashing with Similarity-Adaptive and Discrete Optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP (99): 1-1. DOI: 10.1109/TPAMI.2018.278987.
19. L. Liu, M. Yu and L. Shao, 2017. Latent Structure Preserving Hashing. International Journal of Computer Vision, 122 (3): 439-457.
20. L. Liu, M. Yu and L. Shao, 2015. Projection Bank: From High-dimensional Data to Medium-length Binary Codes. IEEE International Conference on Computer Vision. IEEE, 2821-2829.
21. L. Liu, F. Shen, Y. Shen, X. Liu, L. Shao, 2017. Deep Sketch Hashing: Fast Free-hand Sketch-Based Image Retrieval, IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2862-2871.
22. L. Liu, M. Yu, F. Shen, L. Shao, 2017. Discretely Coding Semantic Rank Orders for Image Hashing. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 5140-5149.
23. C. Yang, D. Han, H. Ko, 2017. Continuous hand gesture recognition based on trajectory shape information. Pattern Recognition Letters, 99: 39-47.
24. W. Wei, Y. Wong, Y. Du, et al. 2017. A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle computer interface, Pattern Recognition Letters, https://doi.org/10.1016/j.patrec.2017.12.005.
25. M. Simao, P. Neto, O. Gibaru, 2017. Using data dimensionality reduction for recognition of incomplete dynamic gestures. Pattern Recognition Letters, 99: 32-38.