Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 3, pp. 1508–1516, DOI 10.12694/scpe.v25i3.2679

A MULTI-LEVEL POWER GRID ENHANCED IDENTITY AUTHENTICATION DATA MANAGEMENT PLATFORM BASED ON FILTERING ALGORITHMS

CAIQIN NONG,[*] HONG ZOU,[†] JIAFA ZHANG,[‡] JIAHAO LIU,[§] PENGFEI SHAO,[¶] JIAWEI JIANG,[‖] AND JIAO XIE[**]

Abstract. In response to the optimal extraction of DCT coefficients in facial images, the author proposes a DCT coefficient extraction method based on discriminant analysis. Based on the discriminant analysis of DCT coefficients, the DCT coefficients with high discriminant values are selected as features. Comparing the DPA based discrete cosine coefficient selection method proposed by the author with the traditional Zigzag discrete cosine coefficient selection method, experiments were conducted on the ORL face database and the Yale face database, respectively. The recognition performance on the ORL face database was higher than that on the Yale face database, as the facial image expression and lighting changes in the ORL database were relatively few, making it suitable for extracting key features. In response to the problem that the speech parameter MFCC is greatly affected by noise and can only reflect the static characteristics of speech, the author extracted gamma pass filtering cepstrum coefficients with human auditory characteristics and gamma pass sliding differential cepstrum coefficients that can reflect the dynamic characteristics of speech based on gamma tone filters and sliding differential cepstrum. In the NUST603 speech database, under pure background, the recognition rate based on GFSDCC features reached 89.88%, and the recognition effect based on GFCC features was 87.52%, which is 4.66% and 2.36% higher than that based on MFCC features. In noisy environments, the average recognition rates of speaker recognition systems based on GFCC and GFSDCC are 56.06% and 59.07%, while the average recognition rates of speaker recognition systems based on MFCC speech features are 53.89%, 2.17% and 5.18% higher, respectively. The gain in this recognition effect comes from the characteristics of the auditory model, as the Gammatone filter effectively reflects the noise resistance of the human auditory system.

Key words: Filtering, Multi level, Enhanced identity authentication, Data management

1. Introduction. The important significance of identity authentication technology in power information systems is reflected in its ability to ensure the security of the power information system, thereby ensuring the security of the entire power system. The current power information system is an important component of power system automation, including numerous automation equipment [10]. Automation equipment has both advantages and disadvantages. The advantage is that it liberates manpower from heavy labor, requiring only necessary monitoring, and plays a very important role in identifying and troubleshooting faults; But at the same time, it also creates a problem where once serious problems occur, the safety and stability of the entire system cannot be guaranteed, thereby affecting the entire power supply work. The current power information system is not perfect, and its identity authentication technology has not been widely applied in the entire industry, resulting in a series of problems. In the future, urban and rural electricity consumption will inevitably increase significantly, and the stable operation of the power system has become an urgent and important responsibility, which should be taken into account in every aspect. Identity authentication refers to the use of various means and methods to identify the identity of a person who wishes to obtain a certain permission. The identification of individuals mostly relies on visual memory, but if machines rely solely on visual recognition, it may result in significant costs [14, 16]. For example, the startup identity authentication of a certain computer can only be done with a simple password, as its security is not as high. If identification devices such as fingerprints and iris are installed on a regular computer, the cost will greatly exceed the budget.There are generally two types of identity

[*]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, Guangdong, 510670, China (Corresponding Author)
[†]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, China
[‡]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, China
[§]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, China
[¶]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, China
[‖]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, China
[**]China Southern Power Grid Digital Grid Group Information and Telecommunication Technology Co., Ltd., Guangzhou, China

authentication technology: Authentication based on relevant information, such as various passwords, certificates, etc; Authenticate based on relevant human characteristics, such as unique elements such as fingerprints and irises. Compared to the two, the former is simple and feasible, and can be relatively freely and widely used. In other words, as long as you master the password or certificate, you can obtain relevant permissions. However, the drawback of this method is also its low security performance, which can cause great inconvenience when relevant passwords and tokens are lost. The latter has the characteristic of uniqueness and is relatively secure, but its limitation is that authenticated users cannot stay away. For example, once users who have entered fingerprints and iris are on a business trip, the system cannot run smoothly. Extracting effective features of faces and speech is the key to completing facial recognition and speech recognition tasks. Although different features can represent facial images and speech signals, they reflect the different characteristics of faces and speech, and their suitable application backgrounds are also different, therefore, how to choose suitable and efficient feature extraction methods based on application needs, and how to improve and improve the performance of existing feature extraction methods are all worth further research. In response to this research issue, Chuang, C. W. et al. used a YOLOv2 model based on deep learning to jointly label the iris and sclera parts of human visible light images, and trained an identity classifier to infer the correct personal identity. The performance of the system was evaluated through a self-made visible light human eye image database, and the average accuracy (mAP) of the proposed iris sclera joint recognition based on deep learning can reach over 99%. In addition, compared to previous work, this design is more effective without the use of iris and sclera segmentation processes [5]. Braeken, A. et al. proposed the first non trivial IBI scheme with implicit authentication using the Elliptic Curve Curved Fanstone (ECQV) implicit authentication scheme. Compared to traditional identity based schemes, implicit certificate based methods can resist key escrow because trusted authorities only have a portion of the keys, which users use as input to construct their own user keys. According to Girault's definition, the scheme can achieve trust level 3 and requires fewer resources compared to certificateless identification. A corresponding formal security model has been defined, demonstrating the resistance of our proposed solution to simulated attacks. Compared with other Schnorr based IBI schemes, our proposed IBI scheme with implicit authentication outperforms other schemes in terms of storage, computing, and communication efficiency, thus providing a feasible solution for applications in the Internet of Things (IoT) environment [2].

Based on current research, the author first proposes a DCT coefficient selection method based on discriminant analysis from the perspective of selecting effective features for facial feature extraction. Secondly, for speech parameter extraction, a Gammatone filter and sliding differential cepstrum are used, We extracted static speech features based on human auditory characteristics, GammatoneFilterCepstralCoefficients (GFCC), and dynamic speech features, GammatoneFilterShiftedDeltaCepstralCoefficients (GFSDCC).

2. Methods.

2.1. Facial DCT feature extraction based on discriminant analysis. Discrete Cosine Transform (DCT) is a common time-domain and frequency-domain transform in signal processing, and has been widely used in feature extraction in face recognition [12]. The DCT transform itself does not perform data compression, it only maps the image source data to another domain. How to select the most effective DCT coefficients as recognition features in the new data domain has become a key issue. The traditional DCT coefficient selection method selects low-frequency DCT coefficients as features in rectangular or Z-shaped order, and the extracted corresponding features often do not represent the best discriminative features. From the perspective of selecting effective features, a DCT coefficient selection method based on Discriminate Power Analysis (DPA) is proposed. Firstly, the DCT coefficients of each position in the facial image are calculated for their discriminative power values, and then the DCT coefficients with higher discriminative power values are selected as feature parameters.

(1) DCT coefficient. For an $M * N$ image matrix $f(x, y)$, its discrete cosine transform is defined as:

$$C(u,v) = a(u)a(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \times \cos\frac{(2x-1)u\pi}{2M} \times \cos\frac{(2y-1)v\pi}{2N} \tag{2.1}$$

Among them, $C(u, v)$ is called the DCT coefficient of matrix $f(x, y), u = 0, 1\ M.\ V = 0, 1\ N.\ a(u)$ and

$a(v)$ are defined as:

$$a(u)=\begin{cases}\sqrt{\frac{1}{2}}, u=0\\ 1,\ \text{otherwise}\end{cases}$$
$$a(v)=\begin{cases}\sqrt{\frac{1}{2}}, v=0\\ 1,\ \text{otherwise}\end{cases} \tag{2.2}$$

After DCT transformation, the two-dimensional DCT coefficients of the image form a matrix of the same size as the original image. The low-frequency coefficients are concentrated in the upper left corner of the matrix, which is the slow-moving part of the image. The high-frequency coefficients are concentrated in the lower right corner of the matrix, which is the detail and edge part of the image [19]. Facial feature extraction has two purposes: Firstly, in order to reduce the dimensionality of the image and the computational complexity during classification, and secondly, in order to select the most representative features to improve classification performance. A large DCT coefficient indicates that the frequency component changes significantly in the facial image. If the coefficient is small, it indicates that the frequency component does not change significantly in the facial image.

The traditional DCT coefficient selection methods, such as the rectangular method, the "Z" shape method, or their related improvement methods, are all based on the decision selection method and do not conduct relevant statistical analysis [9, 18]. Although these feature selection methods are simple and may be effective on certain data samples, they cannot guarantee that they are effective for all samples in the entire database. Based on the different discriminative abilities of each DCT coefficient in the DCT coefficient matrix, the author calculates the discriminative ability value of each DCT coefficient at each position based on discriminative ability analysis, with the aim of selecting DCT coefficients with strong discriminative abilities as features.

(2) Identification ability analysis. The DCT coefficient selection method based on discriminant analysis mainly relies on two assumptions: The coefficient has a large inter class variation and a small intra class variation, which can prove the strong discriminant ability of the coefficient [15]. Assuming the DCT coefficient matrix of a face image of size is:

$$X=\begin{pmatrix} x_{11}, x_{12}, \dots x_{1N} \\ x_{21}, x_{22} \dots x_{2N} \\ \dots \\ x_{M1}, x_{M2} \dots x_{MN} \end{pmatrix} \tag{2.3}$$

Assuming that the training samples have a total of C classes and each class has S images, the training samples have a total of $C*S$ images. Therefore, the calculation process of the discriminant ability value $D(i,j)$ of each DCT coefficient $x_{ij}(i=1,2\cdots M.j=1,2\cdots N)$ in the DCT coefficient matrix can be divided into the following steps:

By selecting the DCT coefficients at positions (i,j) in each DCT coefficient matrix, construct the discriminative ability matrix A_{ij}. The number of matrix A_{ij} is $M*N$, and its form is as follows:

$$A_{ij}=\begin{pmatrix} x_{ij}(1,1), x_{ij}(1,2) \dots x_{ij}(1,C) \\ x_{ij}(2,1), x_{ij}(2,2) \dots x_{ij}(2,C) \\ \dots \\ x_{ij}(S,1), x_{ij}(S,2) \dots x_{ij}(S,C) \end{pmatrix} \tag{2.4}$$

Calculate the average value M_{ij}^{C} for each type of sample:

$$M_{ij}^{C}=\frac{1}{S}\sum_{S=1}^{S} A_{ij}(s,c) \tag{2.5}$$

Calculate the intra class sample mean difference V_{ij}^{C} for each class:

$$V_{ij}^{C}=\sum_{S=1}^{S}\left(A_{ij}(s,c)-M_{ij}^{c}\right)^2 \tag{2.6}$$

Calculate the average value V_{ij}^{W} of the mean difference of samples within Class C:

$$V_{ij}^{W} = \frac{1}{C}\sum_{c=1}^{C} V_{ij}^{c} \tag{2.7}$$

Calculate the average M_{ij} of all samples:

$$M_{ij} = \frac{1}{S}\sum_{C=1}^{C}\sum_{S=1}^{S} A_{ij}(s,c) \tag{2.8}$$

Calculate the sample mean difference V_{ij}^{B} for all samples:

$$V_{ij}^{B} = \sum_{c=1}^{C}\sum_{s=1}^{S}\left(A_{ij}(s,c) - M_{ij}^{C}\right)^2 \tag{2.9}$$

Calculate the discriminative ability value $D(i,j)$ of position (i,j):

$$D(i,j) = \frac{V_{ij}^{B}}{V_{ij}^{W}} \tag{2.10}$$

The larger the discriminant ability value $D(i,j)$, the stronger the discriminant ability value of the DCT coefficient at position (i,j) in the DCT coefficient matrix, indicating that its corresponding DCT coefficient can be selected as a feature parameter [4]. The DCT coefficient selection method based on discriminant analysis is different from previous coefficient selection methods and is a statistical based selection method.

2.2. Static and dynamic speech auditory feature extraction based on Gammatone filter. The human auditory system is an extremely complex perception system. Studying the structure and function of the human ear can not only help us understand the perception process of the human ear, but also greatly assist us in designing automatic processing systems that simulate human ear function. The performance of the human ear auditory system is much more reliable than any automatic speech recognition system [13, 7]. In noisy environments, the speech parameter MFCC is greatly affected by noise and cannot effectively represent speech signals. Moreover, MFCC can only reflect the static characteristics of speech and cannot reflect the dynamic characteristics of speech. The author proposes a Gammatone Filter Cepstral Coefficients (GFCC) based on human ear characteristics based on Gammatone filters. Considering the temporal variation of speech spectrum structure, the author proposes a dynamic parameter for speech, Gammatone Filter Shifted DeltaCepstral Coefficients (GFSDCC), based on the Gammatone filter cepstrum coefficients using sliding differential cepstrum.

(1) Gammatone filtering cepstrum coefficient extraction. MFCC is currently the most commonly used speech feature parameter, among which Mel filter is used to smooth the amplitude square spectrum of speech signals using a triangular filter bank [1]. The author uses Gammatone filter banks instead of Mel filter banks to extract Gammatone filter cepstrum coefficients that can simulate human auditory characteristics. Figure 2.1 shows the extraction process of GFCC. Firstly, the speech signal is preprocessed, followed by Fourier transform of the speech frame to obtain the speech signal spectrum, by using a Gammatone filter, the linear spectral energy is converted into Gammatone spectral energy, and finally its cepstrum value is calculated. That is, the logarithm of the energy is calculated first, and then the discrete cosine transform is performed to obtain GFCC.

(2) Gammatone filtering sliding differential cepstrum. Although GFCC can accurately simulate the auditory characteristics of the human ear and outperform MFCC in recognition performance, like MFCC, it only reflects the static features of speech and does not consider the dynamic characteristics of speech [3]. The differences in people's speech are mainly reflected in the temporal changes in the spectral structure of speech. Shifted Delta Cepstral (SDC) uses a sliding differential cepstrum feature vector composed of several blocks of differential cepstrums spanning multiple frames of speech, allowing one frame feature to contain the acoustic information of multiple frames of speech before and after it, fully reflecting the dynamic characteristics of speech. On the

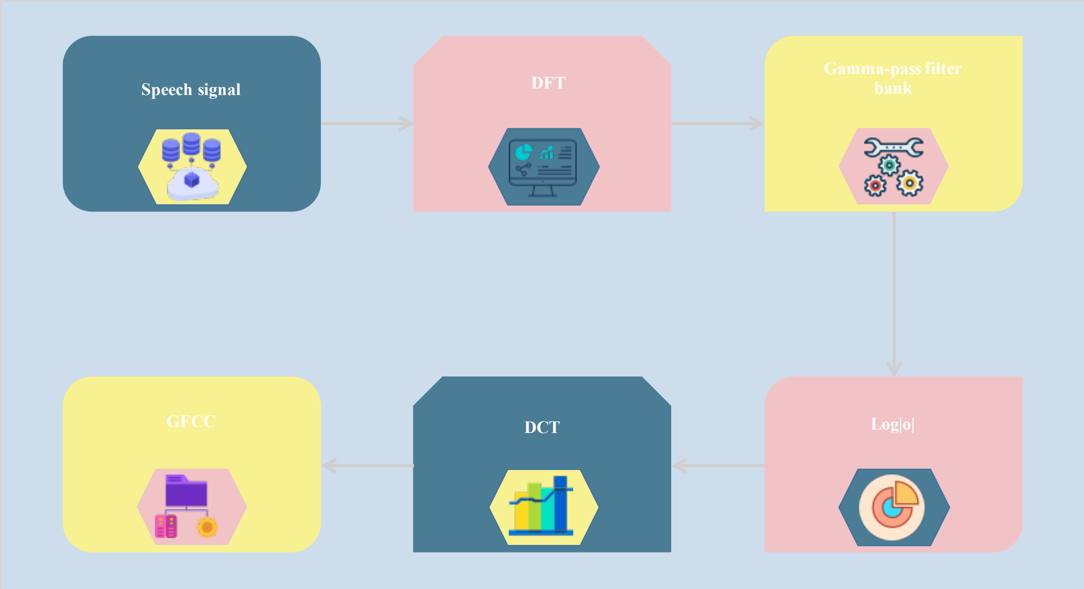

Fig. 2.1: Gammatone filter cepstrum coefficient extraction process

basis of GFCC and SDC, the author proposes a speech feature called Gammatone filtered sliding differential cepstrum coefficient that can reflect the dynamic characteristics of speech and accurately simulate human ear characteristics. By concatenating k-block differential cepstrum, the differential cepstrum is extended within one frame, with each block of differential cepstrum sliding backwards by p frames. During recognition, the author fuses GFCC and GFSDCC in the feature layer to form a fusion feature vector, which not only simulates the auditory characteristics of the human ear, but also comprehensively considers the static and dynamic characteristics of speech features to reduce the impact of external factors on speech signals.

3. Results and Analysis.

3.1. Comparison of different DCT coefficient selection methods. Experimental database: The ORL facial database consists of 400 grayscale facial images from 40 individuals, each with 10 images, and the size of the images is 92 * 112. The background of the image is black, and the facial expressions and details vary, such as whether to smile or not, whether to wear glasses or not, and the facial posture also changes. The depth and plane rotation can reach 20 degrees, and the size of the face can also vary by no more than 10%. In this experiment, each facial image was dimensionally reduced to a size of 46 * 56. The Yale facial database consists of 165 images from 15 individuals, each with 11 facial images, all of which are frontal facial images, with a size of 243 * 320. Facial images have facial expressions, facial details, and changes in lighting. In the experiment, each facial image was dimensionally reduced to a size of 60 * 80.

This experiment will compare the DPA based discrete cosine coefficient selection method proposed by the author with the traditional Zigzag discrete cosine coefficient selection method, and conduct experiments in the ORL face database and Yale face database, respectively. For each individual, the first 5 facial images will be selected as training samples, and the remaining images will be used as test samples. Assuming the training sample is $X_1 = [x_1, x_2, \cdots, x_n]$ and the test sample is $X_2 = [x_1, x_2, \cdots, x_n]$, the Euclidean distance between the two types of samples is as follows:

$$d(X_1, X_2) = \sum^{n} (x_i - x_j)^2 \tag{3.1}$$

Figure 3.1 shows the recognition results of two DCT coefficient selection methods on the ORL face database, while Figure 3.2 shows the recognition results of two DCT coefficient selection methods on the Yale database.

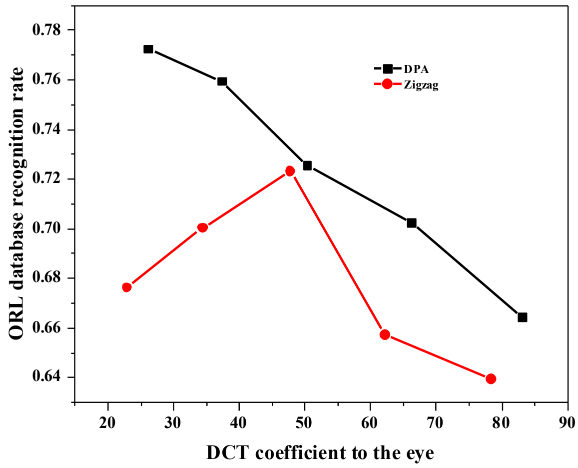

Fig. 3.1: Comparison of recognition rates of different DCT coefficient selection methods on ORL database

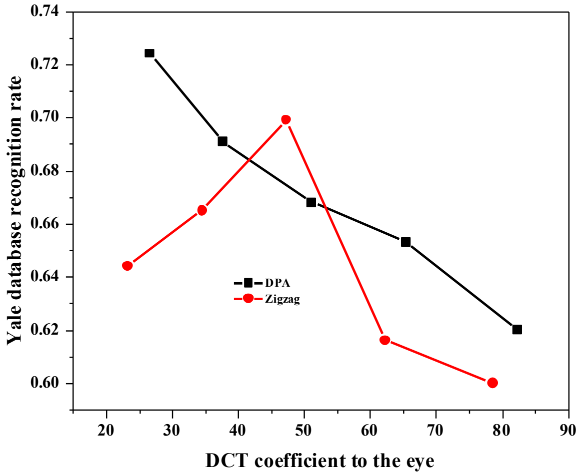

Fig. 3.2: Comparison of recognition rates of different DCT coefficient selection methods on Yale database

From the above two Figures, it can be seen that the DPA method performs better in selecting DCT coefficients than the Zigzag method, regardless of the experimental results in the ORL database or the Yale database. This is because the DCT coefficients selected using the Zigzag method are only low-frequency components in the DCT coefficient matrix, rather than selecting the most discriminative DCT coefficients from the entire coefficient matrix as features for recognition, just like the DPA selection method. From Figures 3.1 and 3.2, it can be seen that the higher the number of DCT coefficients, the higher the recognition rate. When using the DPA method to select DCT coefficients, the highest recognition rate is achieved when selecting 25 coefficients. However, when using the Zigzag method to select DCT coefficients, the best recognition effect is achieved when selecting 49 coefficients. And the recognition performance on the ORL face database is higher than that on the Yale face database, because there are relatively few changes in facial expressions and lighting in the ORL database, which is suitable for extracting key features.

3.2. Comparison of recognition performance of different speech parameters under pure background.

Table 3.1: Comparison of recognition effects of different speech features under pure background

Feature Type	Recognition Rate
MFCC	85.23%
GFCC	87.53%
GFSDCC	89.89%
GFCC+GFSDCC	93.05%

(1) Experimental steps. In order to verify the effectiveness of the speech parameters proposed by the author, the GMM model was used in the NUST603 speech library for validation. The experiment is divided into two parts:

Comparison of recognition performance of different speech parameters under pure background. In a pure background, compare the proposed speech parameters GFCC, GFSDCC, and their fusion features with traditional speech parameters MFCC. Verify the robustness of the speech parameters proposed by the author in different noise environments [6]. Compare the speech parameters GFCC, GFSDCC, and their fusion features proposed by the author with the traditional speech parameter MFCC under the background of White noise and Babble noise with different signal-to-noise ratios.

(2) Experimental database. In this section of the experiment, the author used the NUST603 speech library, which records pure speech in a quiet laboratory environment. The sampling frequency of the speech signal is 22.05KHz, with mono recording and 16Bit quantization. The voice data used in the experiment included 60 speakers, 28 females, and 32 males.

(3) Experimental parameter settings. The speech signal is first preprocessed in the preceding paragraph, using methods based on energy and zero crossing rate for silent detection. Then, a filter with a factor of 0.97 is used for pre emphasis. Then, a frame with a length of 20ms and a frame shift of 10ms is processed, and finally, a Hamming window is processed. Then extract 0-12 dimensional GFCC, totaling 13 dimensions. When extracting GFSDCC features, the selection of its parameter combination N-d-P-k will have a certain impact on the extraction of GFSDCC, among them, N is the number of cepstrum coefficients contained in each frame of speech, d is the time shift for calculating differential cepstrum, p is the sliding frame number of differential cepstrum blocks, and k is the number of differential cepstrum blocks contained in an SDC feature vector. Different parameter combinations have different recognition effects. According to the proposed mountain climbing optimization method, the author adopts a parameter combination of 13-2-3-3, resulting in a total of 39 dimensions of GFSDCC. When fusing the two features in the feature layer, a total of 52 dimension feature vectors are obtained.

In a pure speech environment, test the recognition performance of the GFCC features, GFSDCC features, and their combination features proposed by the author in a GMM model, and compare them with MFCC features. The results are shown in Table 3.1 [8]. From the data in Table 3.1, it can be seen that in pure backgrounds, the recognition performance based on GFCC speech features is better than that based on MFCC speech features, with a recognition rate of 2.3% higher. This is because GFCC features based on Gammatone filter banks have better distinguishability than MFCC features based on Mel filter banks.

GFSDCC features not only utilize the auditory characteristics of Gammatone filter banks, but also incorporate relatively long temporal information into a feature vector, effectively characterizing the dynamic characteristics of language features. In a pure background, the recognition rate based on GFSDCC features reached 89.88%, 2.36% higher than that based on GFCC features, and 4.66% higher than that based on MFCC features [11]. Both GFCC and GFSDCC only reflect one aspect of speech features. The author fused the two in the feature layer, taking into account the static and dynamic characteristics of speech features. In a pure background, the speaker recognition rate based on fused features reached 93.05%, which is 7.82%, 5.52%, and 3.16% higher than the recognition results based on MFCC, GFCC, and GFSDCC, respectively.

(4) Comparison of speech parameter recognition performance in different noise environments. In White noise and Bubble noise environments, under different signal-to-noise ratios, the GMM model was used to test the superiority of the GFCC features, GFSDCC features, and their combination features proposed by the

Table 3.2: Recognition rates using different acoustic features in various speech environments

Feature Type \ Phonetic context		MFCC	GFCC	GFSDCC	GFCC+GFSDCC
	SNR				
Babble noise	0dB	34.73%	35.72%	36.65%	38.96%
	5dB	49.04%	50.41%	53.00%	58.85%
	10dB	52.256%	56.14%	59.49%	63.34%
	15dB	61.59%	64.67%	66.25%	71.25%
	20dB	63.57%	65.33%	68.23%	75.31%
White noise	0dB	36.64%	37.18%	44.02%	51.77%
	5dB	49.78%	50.18%	56.2%	60.73%
	10dB	56.14%	58.05%	61.01%	66.63%
	15dB	65.45%	67.28%	68.72%	73.18%
	20dB	69.8%	75.65%	77.25%	80.1%
average value		53.9%	56.06%	59.08%	64.01%

author in recognition, and they were compared with MFCC features. The results are shown in Table 3.2 [17]. From the experimental results in Table 3.2, it can be seen that in noisy environments, regardless of which Gammatone filter based speech feature is used, the recognition performance is higher than traditional MFCC features. In the Babble noise environment, when the SNR is 0, the recognition rate of the speaker recognition system based on MFCC features is only 34.73%, which is almost impossible to use, the speaker recognition rates based on GFCC and GFSDCC are 35.72% and 36.65%, respectively, while the recognition effect based on the fusion features of the two is 38.96%. In the White noise environment, when the SNR is 20, the recognition rate of the MFCC based speaker recognition system is only 69.8%. The recognition rates of GFCC based and GFSDCC based speakers are 75.65% and 77.25%, respectively, while the recognition effect based on the fusion of the two features is 80.1%. Due to the interference of the "cocktail party" effect, the performance of the speaker recognition system under Babble noise is lower than that under White noise environment. Overall, in a noisy environment, the average recognition rate of a speaker recognition system based on MFCC speech features is 53.9%. The average recognition rates of a speaker recognition system based on GFCC and GFSDCC are 56.07% and 59.08%, while the average recognition rate of a speaker recognition system based on the fusion of the two features is 64.01%, the average recognition rates of speaker recognition systems based on MFCC, GFCC, and GFSDCC speech features are 56.07% and 59.08%, respectively.The average recognition rate of speaker recognition systems based on the fusion of the two features is 64.01%, which is 10.11%, 7.94%, and 4.93% higher than that of speaker recognition systems based on MFCC, GFCC, and GFSDCC speech features, respectively.

From the experimental results in Tables 3.1 and 3.2, it can be seen that the acoustic features GFCC and GFSDCC based on Gammatone filter banks proposed by the author outperform the speech features MFCC in both pure and noisy environments. The gain of this recognition effect comes from the characteristics of the auditory model, as the Gammatone filter well reflects the noise resistance of the human auditory system [20].

4. Conclusion. The author mainly explores the optimal extraction of facial and speech features. Extracting effective features of faces and speech is the key to completing facial recognition and speech recognition tasks. Although different features can represent facial images and speech signals, they reflect the different characteristics of faces and speech, and their suitable application backgrounds are also different, therefore, how to choose suitable and efficient feature extraction methods based on application needs, and how to improve and improve the performance of existing feature extraction methods are all worth further research. The author first addresses the issue of facial feature extraction and proposes a DCT coefficient selection method based on discriminant analysis from the perspective of selecting effective features. After performing DCT transformation on the facial image, the DCT coefficient based discriminant ability values for each position in the image are calculated, and the DCT coefficient with the highest discriminant ability value is extracted as a feature parameter. Secondly, in order to address the issue of speech parameter extraction, static speech features based on human auditory characteristics, GammatoneFilterCepstralCoefficients (GFCC), and dynamic speech features, GammatoneFilterShiftedDeltaCepstralCoefficients (GFSDCC), were extracted using Gammatone filter and sliding differential cepstrum. The acoustic features GFCC and GFSDCC based on Gammatone filter banks proposed by the author outperform speech features MFCC in both pure and noisy environments. The gain of this recognition effect comes from the characteristics of the auditory model, as the Gammatone filter effectively reflects the noise resistance of the human auditory system.

REFERENCES

[1] J. M. Balmer and K. Podnar, *Corporate brand orientation: Identity, internal images, and corporate identification matters*, Journal of Business Research, 134 (2021), pp. 729–737.
[2] A. Braeken, J.-J. Chin, and S.-Y. Tan, *ECQV-IBI: Identity-based identification with implicit certification*, Journal of Information Security and Applications, 63 (2021), p. 103027.
[3] M. L. Charter, *Predictors of feminist identity utilizing an intersectional lens with a focus on non-hispanic white, hispanic, and african american MSW students*, Affilia, 37 (2022), pp. 97–117.
[4] J. Chia, J.-J. Chin, and S.-C. Yip, *A pairing-free identity-based identification scheme with tight security using modified-schnorr signatures*, Symmetry, 13 (2021), p. 1330.
[5] C.-W. Chuang and C.-P. Fan, *Deep-learning based joint iris and sclera recognition with yolo network for identity identification*, Journal of Advances in Information Technology, 12 (2021).
[6] R. Clément and B. Norton, *Ethnolinguistic vitality, identity and power: Investment in SLA*, Journal of Language and Social Psychology, 40 (2021), pp. 154–171.
[7] S. Frandsen and T. Huzzard, *Processes of non-identification: Business school brands and academic faculty*, Scandinavian Journal of Management, 37 (2021), p. 101157.
[8] Q. Gao, C. Zhong, Y. Wang, P. Wang, Z. Yu, and J. Zhang, *Defect analysis of the same batch of substation equipment based on big data analysis algorithm*, 651 (2021), p. 022093.
[9] Y. S. B. C. Z. W. J. S. F. Y. Jiangfeng Zhang, Feiyue Wang, *Research on power grid primary frequency control ability parallel computing based on multi-source data*, Acta Automatica Sinica, 48 (2022), pp. 1493–1503.
[10] X. Li, S. Wang, and Z. Lu, *Reverse identification method of line parameters in distribution network with multi-T nodes based on partial measurement data*, Electric Power Systems Research, 204 (2022), p. 107691.
[11] H. Liu, X. Huizhu, D. Dapeng, L. Fengming, S. Jian, M. Haofei, and W. Qiang, *Topology identification of low-voltage transformer area based on improved particle swarm algorithm*, 1972 (2021), p. 012049.
[12] X. Liu, M. Zhang, X. Xie, L. Zhao, and Q. Sun, *Consensus-based energy management of multi-microgrid: An improved SoC-based power coordinated control method*, Applied Mathematics and Computation, 425 (2022), p. 127086.
[13] O. Matuzkova, I. Rayevska, and O. Grynko, *Identification and identity: Differentiating the conceptual terms*, Wisdom, 1 (2021), pp. 44–52.
[14] F. Peng and J. Zhang, *The broken wires identification of wire rope based on multilevel filtering method using EEMD and wavelet analysis*, Journal of Failure Analysis and Prevention, 21 (2021), pp. 280–289.
[15] L. F. Rowe and M. J. Slater, *Will 'we'continue to exercise? the associations between group identification, identity leadership, and relational identification on group exercise class adherence*, International Journal of Sports Science & Coaching, 16 (2021), pp. 670–681.
[16] J. Song, Y. Jiang, X. Song, Z. Sheng, and Z. Meng, *User-transformer relation identification based on power balance model and adaptive AFSA*, 2195 (2022), p. 012042.
[17] B. Wang, F. Zhao, K. Xu, T. Wen, and L. Jiang, *An efficient multi-parameter synchronous identification method for fiber-reinforced laminated structure based on improved levenberg–marquardt algorithm and modal data*, Journal of Vibration Engineering & Technologies, 11 (2023), pp. 2505–2525.
[18] Y. Wang, W. Zhang, D. Huang, and Y. Liu, *Multi-level feature fusion and multi-loss learning for person re-identification*, Signal Processing: Image Communication, 94 (2021), p. 116197.
[19] L. Yang and T. Huang, *A vehicle reidentification algorithm based on double-channel symmetrical CNN*, Advances in Multimedia, 2021 (2021), pp. 1–6.
[20] Z. Zhang, Z. Wang, H. Wang, H. Zhang, W. Yang, and R. Cao, *Research on bi-level optimized operation strategy of microgrid cluster based on IABC algorithm*, IEEE Access, 9 (2021), pp. 15520–15529.

Edited by: B. Nagaraj M.E.
Special issue on: Deep Learning-Based Advanced Research Trends in Scalable Computing
Received: Oct 8, 2023
Accepted: Nov 25, 2023