Research Article

Multimodal Biometrics Fusion Algorithm Using Deep Reinforcement Learning

Quan Huang (ID)

Institute of Information, Hunan University of Humanities, Science and Technology, Loudi 417000, China

Correspondence should be addressed to Quan Huang; hnldhq@126.com

Multimodal biometrics fusion plays an important role in the field of biometrics. Therefore, this paper presents a multimodal biometrics fusion algorithm using deep reinforcement learning. In order to reduce the influence of user behavior, user's personal characteristics, and environmental light on image data quality, data preprocessing is realized through data transformation and single-mode biometric image region segmentation. A two-dimensional Gobar filter was used to analyze the texture of local sub-blocks, qualitatively describe the similarity between the filter and the sub-blocks and extract the phase information and local amplitude information of multimodal biometrics features. Deep reinforcement learning was used to construct the classifier of different modal biometrics, and the weighted sum fusion of different modal biometrics was implemented by fractional information. The multimodal biometrics fusion algorithm was designed. The Casia-iris-interval-v4 and NFBS datasets were used to test the performance of the proposed algorithm. The results show that the fused image quality is better, the feature extraction accuracy is between 84% and 93%, the average accuracy of feature classification is 97%, the multimodal biometric classification time is only 110 ms, the multimodal biometric fusion time is only 550 ms, the effect is good, and the practicability is strong.

1. Introduction

With the development of biometric extraction and its related technologies, biometric extraction products are becoming more and more mature and are gradually entering every aspect of society, which has brought great changes in people's living habits. Among them, fingerprint identification attendance meter, face recognition access control, iris customs clearance system, and other products have played a great role in security and brought great convenience to people's lives. However, with the increasing demand for security, there is an urgent need for a better biometric recognition system that must meet various requirements, including uniqueness, universality, stability, collectability, ease of use, and security. Universality means that the attribute of use needs to be the attribute of normal and healthy human beings; uniqueness means that the biometric information used is unique to the individual; stability means that the biometric mode needs to be stable and constant for a long time; collectability means that biometric modal

information can be collected by external equipment; and ease of use means that the system needs to meet the needs of users for convenience [1]. Security means that the mode needs to be not easy to leave so as to avoid theft and forgery by others. The above requirements can be met through multimodal biometric fusion. Research on this problem plays a positive role in the improvement of research results in the field of biometric extraction and has high research value. Therefore, the multimodal biometric fusion algorithm is studied [2].

With more and more attention paid to security in various fields, research on multimodal biomedical feature fusion algorithms has gradually attracted much attention. Dinh Phu-Hung [3] proposed a multimodal medical image fusion algorithm based on the marine predator algorithm and three-scale image decomposition. The algorithm uses the Kirsch compass operator based on the local energy function for detailed level fusion so that the output image retains important information. Moreover, the algorithm uses the marine predator algorithm (MPA) to fuse the

bottom layer by optimizing the parameters so that the output image has good quality, but the algorithm has the problem of low accuracy in feature extraction. Xinhua Li and Jing Zhao [4] proposed a new algorithm for multimodal medical image fusion. Firstly, CT images and MRI images are decomposed into low and high-frequency sub-bands by NSCT with multiscale geometric transformation; secondly, for the low-frequency sub-band, the local standard deviation algorithm or fusion is selected. For the high-frequency sub-band, the adaptive pulse-coupled neural network model is constructed, and the fusion rules are set according to the cumulative ignition times of iterative operations in the network; finally, the fused image is obtained by image reconstruction. However, the algorithm has the problem of a low error rate. M. Zhu and X. Yu [5] proposed a multifeature fusion algorithm in detail enhancement of VR panoramic images. The shadow detection results are obtained by using HSV color features and texture features, and then the final detection results are obtained by fusion. Experimental results show that the algorithm greatly reduces the false detection rate. However, the algorithm has the problem of low average classification accuracy. Geng Peng and Xiuming Sun [6] proposed a multimodal medical image fusion algorithm based on the quaternion wavelet transform. The algorithm can fuse not only CT and MR image pairs, but also CT and proton density-weighted MR image pairs, as well as multispectral MR images such as T1 and T2. However, the algorithm has the problem of a high average authentication time for a single user. Castro et al. [7] proposed a multimodal feature fusion algorithm for gait recognition based on CNN. Based on the original pixels and their derived simple functions, the algorithm uses advanced learning technology to extract relevant features and fuse the original pixel information with the information from optical flow and depth map to realize multimodal feature fusion. However, the algorithm has the problem of a low error rate. Therefore, a new multimodal biometrics fusion algorithm based on deep reinforcement learning is proposed. The main contributions to this paper are as follows: (1) data preprocessing is realized through data transformation, single-mode biometric image region segmentation, and other steps to reduce the impact of user behavior, user personal characteristics, and ambient light on image data quality. (2) The two-dimensional Gobar filter is used to analyze the local sub-blocks of texture, qualitatively describe the similarity between the filter and the sub-blocks, extract the phase information and local amplitude information of multimodal biometrics, and the

accuracy and efficiency of feature extraction are proposed. (3) Deep reinforcement learning is used to construct different modal biometric classifiers, and the weighted sum fusion of different modal biometrics is implemented through the score information to realize the design of a multimodal biometric fusion algorithm so as to achieve the multiple objectives of improving the fusion efficiency and quality.

2. Materials and Methods

2.1. Dataset Description. In the process of testing the designed multimodal biomedical feature fusion algorithm based on deep reinforcement learning, the experimental datasets used are the CASIA-Iris-Interval-v4 and NFBS datasets, respectively.

The CASIA-Iris-Interval-v4 includes binocular iris images of 100 users. The image size is $640 * 480$, and the acquisition times on the left and right sides of each person are 7 times, for a total of 1400 images. The NFBS (neuro-feedback skull-stripped) dataset, the dataset includes 125 sets of brain image data. Each set of data includes a defaced T1 MR image, a skull removed MR image, brain segmentation results, and its dataset modal T1 weighted MRI scan. In the performance test of the algorithm, the experimental parameters are set as follows: the corresponding input batch size of a convolutional neural network is set to 16. Adam is the optimizer; the initial learning rate is set to 0.01; the probability of random inactivation of neurons is set to 0.8. The two datasets are divided into verification set, test set, and training set, with a division ratio of $1:2:7$.

2.2. Preprocessing of Single-Modal Biometric Image Data. Preprocess the single-modal biometric image data to eliminate the adverse effects of user behavior, user personal characteristics, and ambient light. The specific data preprocessing steps are data transformation and single-modal biometric image region segmentation.

Data transformation is to transform the value range of single-modal biometric image data according to demand so as to better implement the processing of single-modal biometric image data [8]. The processing method of data transformation is the minimum-maximum normalization method, which mainly implements the linear transformation of value range data through the linear transformation method [9] so that the transformed data is in a unified value range. The equation is as follows:

$$new\_value = \frac{old\_value - old\_Min \times (new\_Max - new\_Min)}{old\_Max - old\_Min} + new\_Min, \qquad (1)$$

where new_value refers to the sample data within the normalized single-modal biometric range; old_value refers to the sample data of single-modal biometric range; old_Min refers to the minimum value of sample data in value range; new_Max refers to the maximum

value of sample data in value range; new_Min refers to the minimum value of the data classification range within the value range; and old_Max refers to the maximum value of the data classification range within the value range [10].

The single-modal biometric region segmentation mainly detects the edge between the background region and the single-modal biometric region; that is, the single-modal biometric contour is detected, and the single-modal biometric region is segmented in the image background [11].

The human body is affected by infrared, and the brightness of the single-modal biometric region is often lower than that of other background regions. Edge detection is implemented by the Canny operator. The detection steps are as follows:

(1) The smoothing filter is constructed by a one-dimensional Gaussian function $G(x)$. The convolution operation of $f(x, y)$ is performed by column and row, respectively, to obtain the smooth image $I(x, y)$. The equation for constructing the filter is as follows [12]:

$$G(x) = \frac{\exp\left(-x^2/2\sigma^2\right)}{2\pi\sigma^2}, \tag{2}$$

  (i) where σ is the standard deviation corresponding to the Gaussian function.

(2) We pass a 2×2 matrix, where the gradient direction $H(x, y)$ and gradient amplitude $M(x, y)$ of $I(x, y)$ are calculated by the neighborhood first-order partial derivative equation [13]. The equation is as follows:

$$
\begin{aligned}
H(x, y) &= \arctan\left[k_x(x, y), k_y(x, y)\right], \\
M(x, y) &= \sqrt{k_x^2(x, y) + k_y^2(x, y)},
\end{aligned}
\tag{3}
$$

  (i) where k_x refers to the result of $I(x, y)$ convolution operation of the smooth image by line through the filter; k_y refers to the result of $I(x, y)$ convolution operation of the smooth image by column through the filter. The convolution template of k_x is as follows:

$$f_x = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}. \tag{4}$$

  (ii) The convolution template of k_y is as follows:

$$f_y = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}. \tag{5}$$

(3) The nonmaximum suppression of gradient amplitude is implemented to obtain all possible edge points in the smooth image $I(x, y)$.

(4) The result of step (3) is segmented by high and low thresholds, respectively, to obtain two threshold edge images. For the image obtained by low threshold segmentation, the edges need to be collected continuously until all the gap connections of the image obtained by high threshold segmentation are realized [14].

Through the above steps, the preprocessing of single-modal biometric image data is completed.

2.3. Multimodal Biometric Extraction.
The texture local sub-block analysis of the segmented single-modal biometric region is implemented through the two-dimensional Gobar filter. The similarity between the filter and the local sub-block is qualitatively described [15], and the phase information and local amplitude information of multimodal biometrics are extracted [16]. When the convolution result is positive, the local texture phase coding of the sub-block is 1. On the contrary, the local texture phase coding is 0.

The definition of a two-dimensional Gobar filter is as follows:

$$R(x, y) = e^{\pi \left[\frac{(x - x_0)^2 + (y - y_0)^2}{\lambda \cdot \tau} \right]} e^{\pi^2 \left[\frac{(x - x_0)^2 + (y - y_0)^2}{\lambda_0 \cdot \tau_0} \right]}, \tag{6}$$

where x_0 refers to the time domain frequency of two-dimensional Gabor wavelet; y_0 refers to the frequency of two-dimensional Gabor wavelet in the frequency domain; λ refers to the localization frequency of two-dimensional Gabor wavelet; τ_0 refers to the local texture threshold of sub-block; λ_0 refers to the wavelet iterative function; and τ_0 refers to the scale difference function of texture extreme points.

The texture coordinates in the sub-block are transformed into a fixed scale rectangular region to describe each pixel in the single-modal biometric region.

$$R(r, \theta) = e^{-\omega(\theta - \theta')} e^{\left(-(r - r')^2/\lambda^2\right)} e^{\pi \left(-(\theta - \theta')^2/\tau^2\right)}, \tag{7}$$

where r refers to coordinate transformation distance; θ refers to coordinate transformation angle; ω refers to rectangular area's scale; r' refers to the normalized value of coordinate transformation distance; and θ' refers to the normalized value of coordinate transformation angle [17, 18].

Through $R(x, y)$, filter waves of $R(r, \theta)$. The local sub-block is qualitatively described as 1 and 0. The single-modal biometrics are described through the description of the local sub-block, and the corresponding multibit feature codes of various single-modal are created to realize the extraction of multimodal biometrics [19, 20].

2.4. Multimodal Biological Feature Fusion Algorithm.
Deep reinforcement learning is used to construct different modal biometric classifiers, and the weighted sum fusion of different modal biometric features is implemented through the score information to realize the design of a multimodal biometric fusion algorithm.

  (i) Input: test sample set of multimodal biometrics.

  (ii) Output: fusion results of multimodal biometrics.

  (1) The deep reinforcement learning algorithm based on convolution neural networks is used to construct different modal biometric classifiers. The classifier is composed of two modules: Inception-ResNet module and residual connection module [21].

The residual connection module can accumulate the output result of the upper layer network and the copy data and directly input it to the lower layer network. The accumulation of the output and the original image feature map helps to reduce the training error. The module network is composed of multiple residual blocks. It is a stacked deepening network that can prevent the occurrence of the overfitting phenomenon. The middle of the network is the residual structure, which can realize jump transmission [22, 23].

The Inception-ResNet module uses the inception structure and can find the optimal sparse structure unit. The single-modal between the two activation functions is expanded to become a new neural network, and multiple convolution kernels of different sizes are used to realize feature classification on different scales. The loss function of the classifier is as follows:

$$l = \frac{1}{2} \sum_{i=1}^{N} |t - C_i|^2, \tag{8}$$

where l refers to the loss value; N refers to the number of samples; t refers to the feature value; and C_i refers to the feature center point of category corresponding to sample i.

When the value of the loss function is the smallest, the network can reach the most convergent state.

(2) m classifiers are represented by the following equation:

$$M = \{m_1, m_2, \ldots, m_n\}, \tag{9}$$

where m_n represents the n th classifier.

The classification results of m classifiers are as follows:

$$H = \sum_{n=1}^{M} \alpha m_n, \tag{10}$$

where α represents the average classification threshold:

(3) Let us estimate the optimal weight of classification results as follows:

$$\delta_n = \left(\frac{\kappa(\mathrm{Gen}_n) - S}{\kappa(\mathrm{Imp}_n) - S} \right)_2^2, \tag{11}$$

where $\kappa(Gen_n)$ is the mean value of the truth function $\kappa(\mathrm{Imp}_n)$ is the mean of false value scores. S is the mean of the matching score.

(4) Let us quantify the uncertain area in the feature information through the mean closure strategy as follows:

$$\mu = F(z_i)S, \tag{12}$$

where μ refers to the quantized value of uncertain area and $F(z_i)$ refers to the quantized function of uncertain value z_i.

(5) Let us assign weight to uncertain areas according to δ_n value as follows:

$$s_i = \frac{\delta_n}{\sum_{n=1}^{M} \delta_n}. \tag{13}$$

The assigned weight value satisfies the following equation:

$$0 \leq s_i \leq 1, \quad \forall M,$$
$$\sum_{n=1}^{M} s_i = 1, \quad \forall n. \tag{14}$$

Through the reasonable distribution of weights, the summation and weighted fusion of the matching score information of different modal features are realized. The process of the proposed algorithm is shown in Figure 1.

2.5. Experimental Index. In the experiment, multimodal medical image fusion algorithms based on marine predator algorithm and three-scale image decomposition, the new algorithm for multimodal medical image fusion, multi-feature fusion algorithm in detail enhancement of VR panoramic images, multimodal medical image fusion algorithm based on quaternion wavelet transform, and multimodal feature fusion algorithm for gait recognition based on CNN are used as the comparison test algorithms, which are represented by algorithms 1, 2, 3, 4 and 5, respectively.

The various performances of the proposed algorithm are tested. The experimental indexes are as follows: fusion effect: we take brain CT images and MRI images as examples for multimodal biometric fusion, and compare the fusion results, that is, the quality of the image. Accuracy of multimodal biometric extraction: the calculation equation of multimodal biometric extraction accuracy is as follows:

$$A = \frac{a_i}{a_j} \times 100\%, \tag{15}$$

where a_i refers to the amount of data correctly extracted from multimodal biometric samples, and a_j refers to the total number of samples for multimodal biometrics.

Average classification accuracy (ACA): it represents the average probability of correct classification of multimodal biometrics in multiple experiments, and its calculation equation is as follows:

$$\mathrm{ACA} = \frac{\sum_{i=1}^{n} \left((b_{ik}/b_{ij}) \times 100\% \right)}{n}, \tag{16}$$

where a_{ik} refers to sample data size of correctly classified multimodal biometrics, b_{ij} refers to the experimental data, and n refers to the times of experiments.

The time-consuming of multimodal biometric classification: the time-consuming of multimodal biometric classification refers to the time spent in classifying multimodal biometrics. The calculation equation of the index is as follows:

$$T_1 = |t_1 - t_2|, \tag{17}$$

where t_1 represents the classification start time and t_2 represents the classification end time.
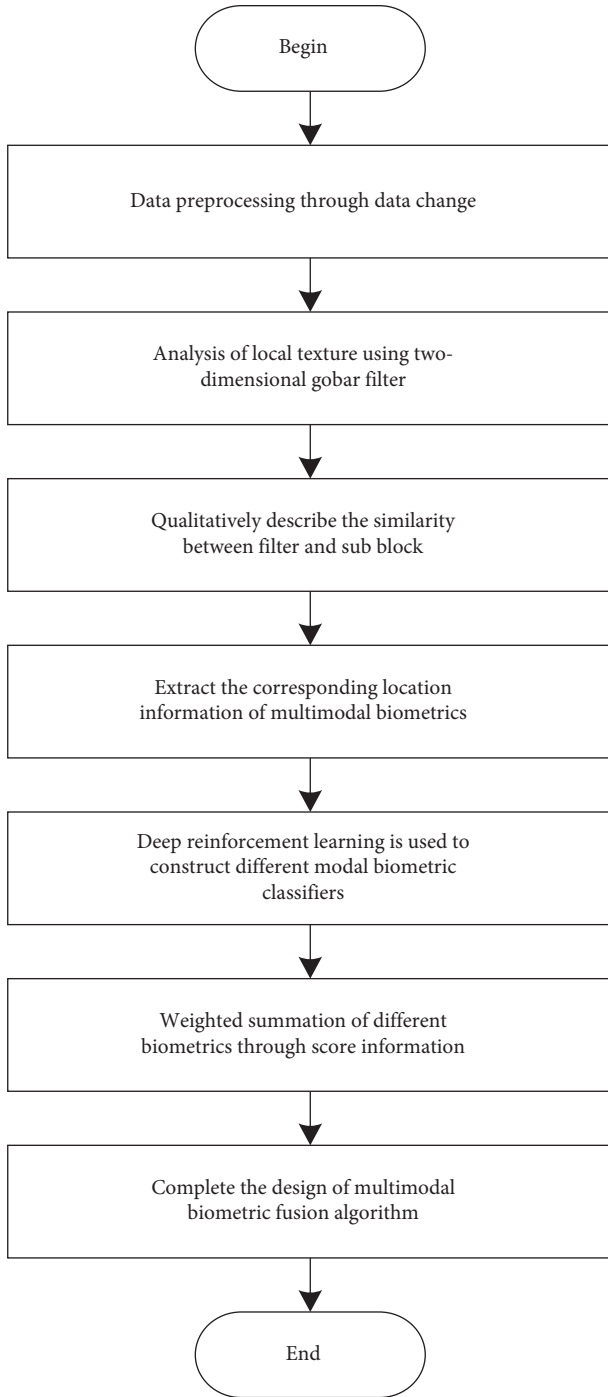
FIGURE 1: Process of multimodal biometric fusion algorithm.

Multimodal biometric fusion time: it refers to the sum of the time taken to complete all multimodal biometric fusion steps. It is expressed as follows:

$$T_2 = \sum_{i=1}^{n} t_i, \quad (18)$$

where t_i represents the time taken for the i th multimodal biometric fusion step.

3. Results and Discussion

This paper takes brain CT images and MRI images as an example for multimodal biometric fusion, in which the original image is shown in Figure 2.

The fusion results of different algorithms are shown in Figure 3.

The fusion results of brain CT images and MRI images are shown in Figure 3. The fusion images with algorithms 1–5 are of low quality and have poor definition. Therefore, the fusion results cannot accurately reflect the detailed features of the two images. Compared with these algorithms, the proposed algorithm almost retains the texture of the soft tissue area in most brain MRI images and obviously retains the shadow area of a suspected blood clot in brain CT images. The image definition is higher and the fusion effect is better.

We test the feature extraction accuracy of the designed multimodal biometric fusion algorithm based on deep reinforcement learning and five comparison algorithms. The test results are shown in Figure 4.

According to the data in Figure 4, the process of increasing dimension, the feature extraction accuracy of the proposed algorithm also increases. The feature extraction accuracy of algorithm1 is 74%–86%, that of algorithm 2 is 63%–73%, that of algorithm 3 is 52%–64%, that of algorithm 4 is 57%–70%, and that of algorithm 5 is 56%–67%. The feature extraction accuracy of the proposed algorithm is between 84% and 93%, and its feature extraction accuracy is higher than that of the five comparison algorithms, which shows that the proposed algorithm can achieve more accurate multimodal biometric extraction.

We test the average classification accuracy of the designed algorithm and the five comparison algorithms. The test results are shown in Figure 5.

The average classification accuracy test results in Figure 5 show that the highest average classification accuracy was 82% for algorithm 1, 73% for algorithm 2, 76% for algorithm 3, 72% for algorithm 4, and 75% for algorithm 5. The average classification accuracy of multimodal biometrics of the proposed algorithm is high, up to 97%, which is higher than the five comparison algorithms, indicating that the ACA performance of multimodal biometrics of the proposed algorithm is strong.

The time-consuming test results of multimodal biometric classification are shown in Figure 6.

The test results in Figure 6 show that the multimodal biometric classification of algorithm 1 takes 270 s, the multimodal biometric classification of algorithm 2 takes 340 s, the multimodal biometric classification of algorithm 3 takes 250 s, the multimodal biometric classification of algorithm 4 takes 260 s, and the multimodal biometric classification of algorithm 5 takes 460 s. The multimodal biometric classification of the proposed algorithm takes the shortest time, and it only takes 110 ms to realize the multimodal biometric classification with the highest efficiency.

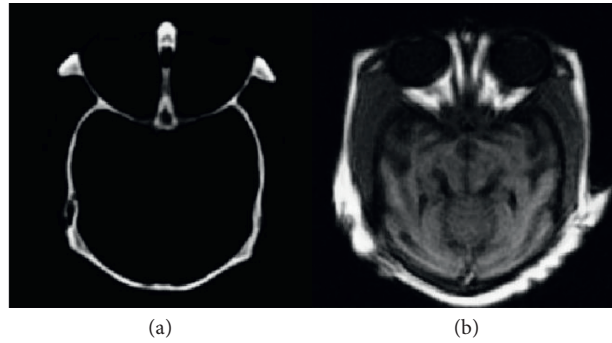The test results of multimodal biometric fusion time of six algorithms are shown in Figure 7.

(a)   (b)

FIGURE 2: Original images: (a) CT image and (b) Mr-t2 image.

(a)   (b)   (c)

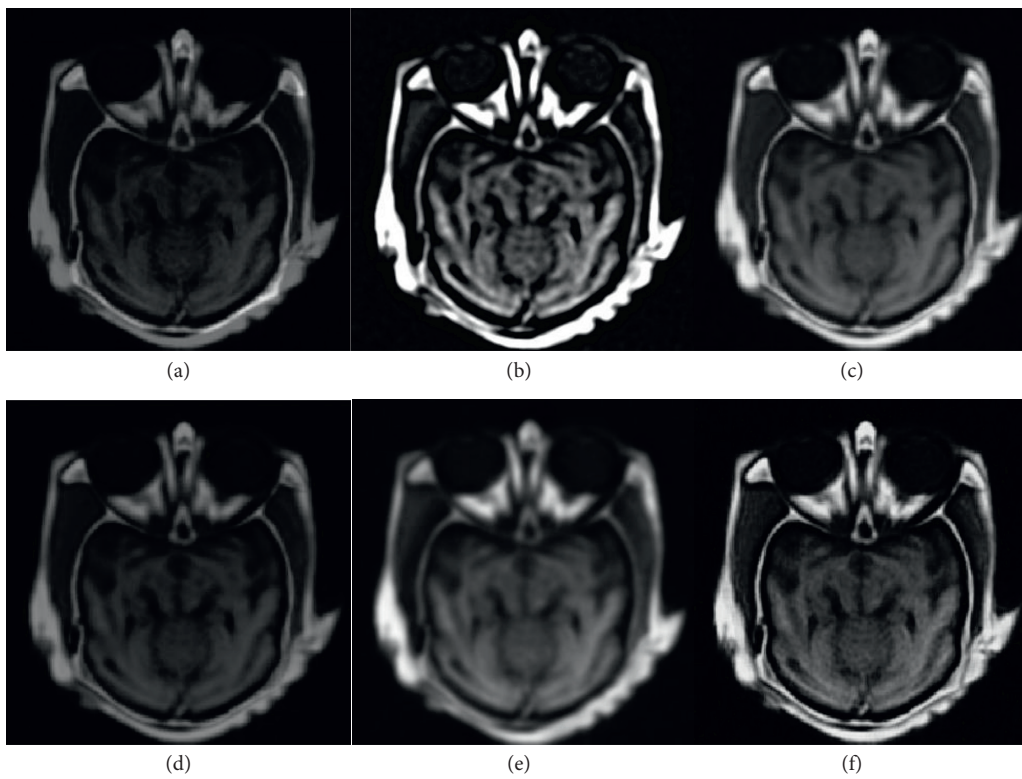(d)   (e)   (f)

FIGURE 3: Comparison of fusion effects: (a) algorithm 1, (b) algorithm 2, (c) algorithm 3, (d) algorithm 4, (e) algorithm 5, and (f) the proposed algorithm.

According to the multimodal biometric fusion time test data in Figure 7, the multimodal biometric fusion time of algorithm 1 is 700 ms, the multimodal biometric fusion time of algorithm 2 is 740 ms, the multimodal biometric fusion time of algorithm 3 is 640 ms, the multimodal biometric fusion time of algorithm 4 is 700 ms, and the multimodal biometric fusion time of algorithm 5 is 650 ms. The multimodal biometric fusion time of the proposed algorithm is the shortest, and the multimodal biometric fusion can be realized in only 550 ms, which is more efficient.
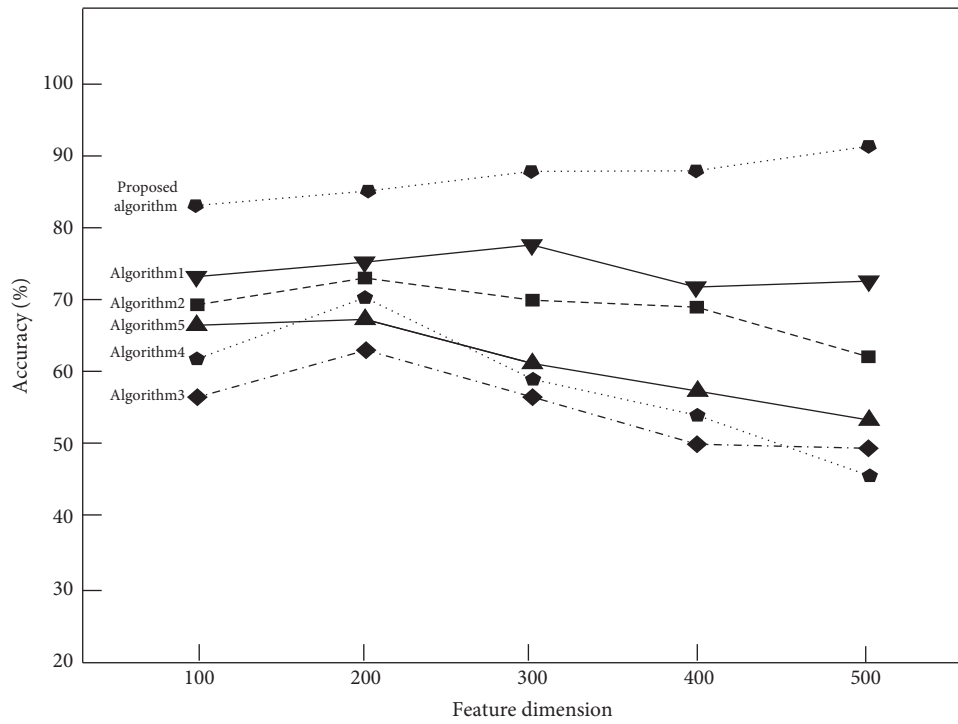
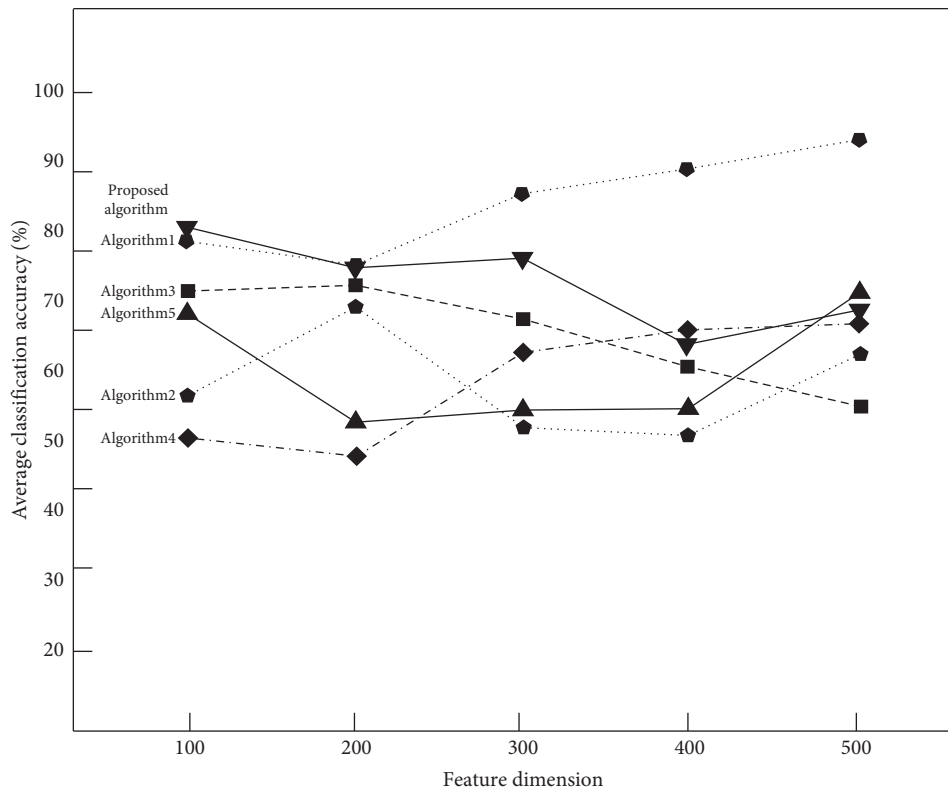FIGURE 4: Comparison of feature extraction accuracy results.

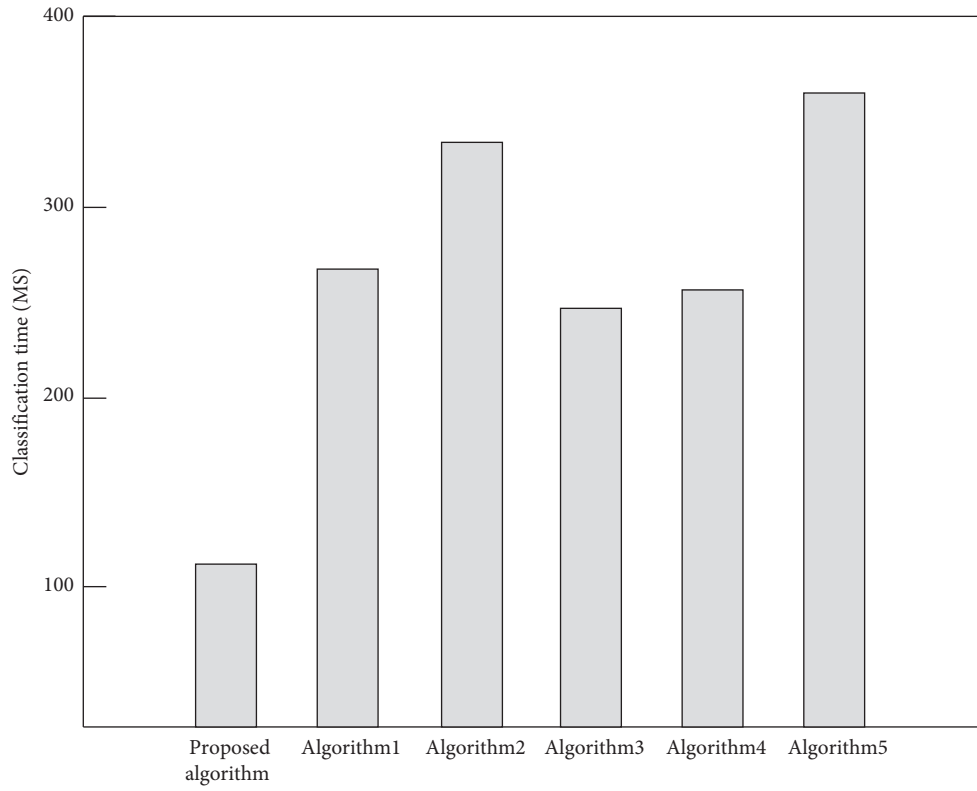FIGURE 5: Comparison of average classification accuracy results.

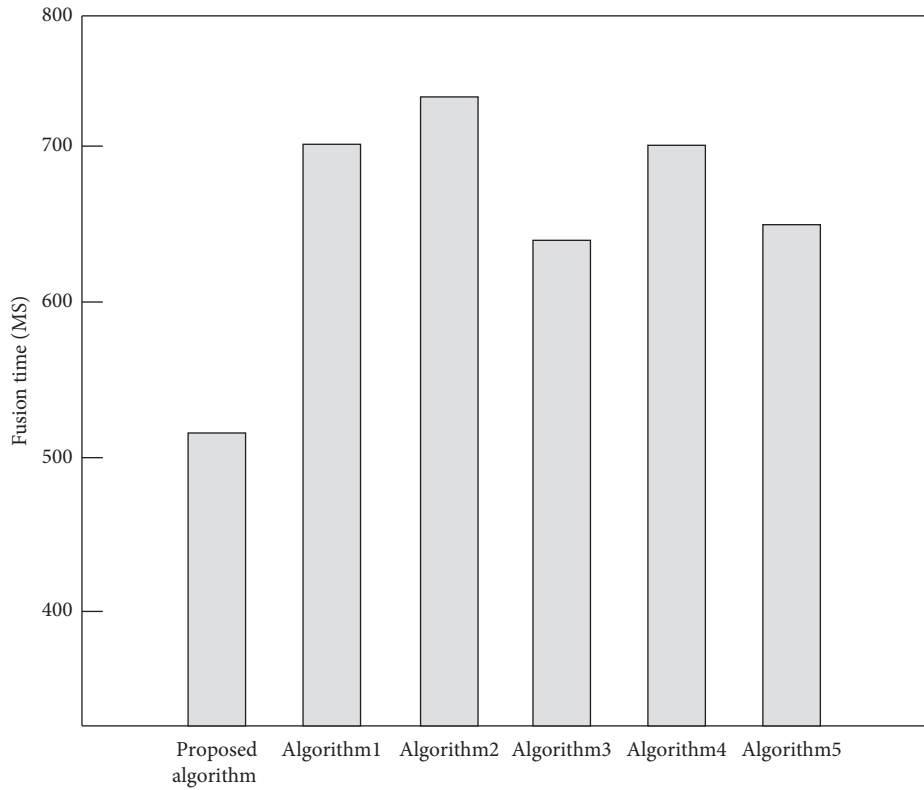FIGURE 6: Comparison of classification time-consuming test results.

FIGURE 7: Comparison of fusion time test results.

4. Conclusions

In the research, the problem of multimodal biometric fusion is solved, a multimodal biometric fusion algorithm based on deep reinforcement learning is proposed, and the algorithm is verified in practice. The following research results were obtained: (1) the unification of the value range of single-modal biometric image data and the segmentation of single-modal biometric regions are realized; (2) the extraction of multimodal biometric phase information and local amplitude information is realized by two-dimensional Gobar filter; and (3) different modal biometric classifiers are designed. In the research, due to the limitations of time and energy, the training of classifiers and other issues have not been studied in detail. More in-depth and detailed research will be carried out in the future so as to maximize the classification accuracy and the effect of multimodal biometric fusion.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest with this manuscript.

Acknowledgments

References

[1] G. S. Walia, G. Jain, N. Bansal, and K. Singh, "Adaptive weighted graph approach to generate multimodal cancelable biometric templates," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1945–1958, 2020.

[2] D. Zhao, Z. Chang, and S. Guo, "A Multimodal fusion approach for image captioning," *Neurocomputing*, vol. 329, pp. 476–485, 2019.

[3] P. H Dinh, "A novel approach based on Three-scale image decomposition and Marine predators algorithm for multimodal medical image fusion," *Biomedical Signal Processing and Control*, vol. 67, no. 2, pp. 1–14, 2021.

[4] X. Li and J. Zhao, "A novel multimodal medical image fusion algorithm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 1–8, 2020.

[5] M. Zhu and X. Yu, "Multi-feature fusion algorithm in VR panoramic image detail enhancement processing," *IEEE Access*, no. 99, p. 1, 2020.

[6] G. Peng, X. Sun, and J. Liu, "Adopting quaternion wavelet transform to fuse multimodal medical images," *Journal of Medical and Biological Engineering*, vol. 37, no. 5, pp. 230–239, 2017.

[7] F. M Castro, M. J Marín-Jiménez, N. Guil, and N. Blanca, "Multimodal feature fusion for CNN-based gait recognition: an empirical comparison," *Neural Computing & Applications*, vol. 32, no. 11, pp. 14173–14193, 2020.

[8] C. Taouche and H. Belhadef, "Multimodal biometric system combining left and right palmprints," *Information Discovery and Delivery*, vol. 48, no. 1, pp. 2–13, 2019.

[9] D. Zhao, "Rapid multimodal image registration based on the local edge histogram," *Mathematical Problems in Engineering*, vol. 2021, no. 9, pp. 1–9, 2021.

[10] Z. Wang, J. Zhen, Y. Li, G. Li, and Q. Han, "Multi-feature multimodal biometric recognition based on quaternion locality preserving projection," *Chinese Journal of Electronics*, vol. 28, no. 4, pp. 789–796, 2019.

[11] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. A. El-Samie, "Hybrid pixel-feature fusion system for Multimodal medical images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 42, pp. 1–18, 2021.

[12] D. Wang, Y. Li, L. Jia, Y. Song, and Y. Liu, "Novel three-stage feature fusion method of multimodal data for bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 99, pp. 1–10, 2021.

[13] S. Bhagat, S. D. Joshi, B. Lall, and G. Smriti, "Multimodal sensor fusion using symmetric skip autoencoder via an adversarial regulariser," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, p. 1, 2020.

[14] L. C. O. Tiong, S. T. Kim, and Y. M. Ro, "Implementation of Multimodal biometric recognition via multi-feature deep learning networks and feature fusion," *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22743–22772, 2019.

[15] C. Zhang, "Convolution analysis operator for Multimodal image fusion," *Procedia Computer Science*, vol. 183, no. 5, pp. 603–608, 2021.

[16] L.-W. Ko, C.-T. Lin, Y.-C. Lu et al., "Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface," *IEEE Computational Intelligence Magazine*, vol. 14, no. 1, pp. 96–106, 2019.

[17] X. Gui, X. Liu, Q. Tian, and W. Guo, "Multimodal Data Fusion in 3-D Printing Quality Prediction," *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[18] N. Bakalos, A. Voulodimos, N. Doulamis et al., "Protecting water infrastructure from cyber and physical threats: using multimodal data fusion and adaptive deep learning to monitor critical systems," *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 36–48, 2019.

[19] A. Shamsan, W. Dan, and C. Cheng, "Multimodal data fusion using multivariate empirical mode decomposition for automatic process monitoring," *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[20] Y. R. Pandeya and J. Lee, "Deep learning-based late fusion of Multimodal information for emotion classification of music video," *Multimedia Tools and Applications*, vol. 80, no. 38, pp. 1–19, 2021.

[21] Z. Wu, K. Mao, and G.-W. Ng, "Enhanced feature fusion through irrelevant redundancy elimination in intra-class and extra-class discriminative correlation analysis," *Neurocomputing*, vol. 335, pp. 105–118, 2019.

[22] Y. Wu, Y. Zhao, and X. Lu, "Modeling incongruity between modalities for multimodal sarcasm detection," *IEEE Multimedia*, vol. 28, no. 99, p. 1, 2021.

[23] H. Mehraj and A. H. Mir, "A multi-biometric system based on multi-level hybrid feature fusion," *Herald of the Russian Academy of Sciences*, vol. 91, no. 2, pp. 176–196, 2021.