DOI: 10.19884/j.1672-5220.202501008

Human Gait Recognition Based on SOM-K-Means Clustering for Lunar Landing Spacesuit

HAO Boda, LÜ Hongzhan*
College of Mechanical Engineering, Donghua University, Shanghai 201620, China

Abstract: Whether the lunar landing spacesuit can accurately recognize the astronaut's gait and provide timely buffering upon lower limb impact is a critical factor in ensuring the safety of extravehicular operations. This study addresses this issue by segmenting the human gait phases and collecting gait data in a simulated lunar surface low-gravity environment. Z-score normalization and Gaussian filtering are applied for data preprocessing. The self-organizing map (SOM)-K-means clustering algorithm is adopted to establish the gait recognition model, and the model is then evaluated. Experimental results and comparisons with other clustering algorithms validate the effectiveness of the SOM-K-means algorithm for gait recognition and the generalizability of the model. The results demonstrate that the model achieves satisfactory clustering performance and accurate gait recognition, providing technical support for the design of high-performance lunar landing spacesuits.

Keywords: lunar landing spacesuit; gait recognition; self-organizing map (SOM); K-means clustering
CLC number: TP391; V445.3 **Document code**: A
Article ID: 1672-5220(2026)04-0102-08

0 Introduction

The primary mode of manned lunar exploration involves astronauts conducting extravehicular activities on the lunar surface. The "bouncing" gait exhibited by astronauts when walking on the moon is the most significant form of locomotion. However, this behavior can easily cause impact damage to astronauts' lower limb bones. Therefore, lunar landing spacesuits must be capable of buffering the impact forces transmitted to astronauts' lower limbs. Consequently, it is essential for these spacesuits to accurately identify the astronauts' walking gait.

Human gait recognition encompasses the identification of gait events, phases, spatiotemporal parameters, and types[1]. The importance of gait recognition algorithms has been widely recognized, with extensive research conducted across fields such as medicine, industry, daily living, and military applications.

As an interdisciplinary field combining biomechanics and pattern recognition, human gait recognition has evolved toward multimodal fusion in engineering systems. Computer vision-based methods show significant advantages in open environments. Ryu et al.[2] developed a spherical space model using human point clouds for recognizing frontal-view human gaits, achieving high accuracy. However, its dependence on a fixed viewpoint limited its applicability in spacesuit multi-degree-of-freedom motion scenarios. Han et al.[3] proposed the gait energy image (GEI) to characterize human gait and designed a recognition method combining real and synthetic templates. Although GEI effectively captures gait cycle features, it suffers from feature drift when dealing with gait variability caused by unstructured lunar surface terrain. Deng et al.[4] used gait features acquired with Kinect sensors, combining deterministic gait features with deterministic learning theory and Kinect data streams to improve dynamic adaptability. However, the red-green-blue-depth (RGB-D) sensor's signal-to-noise ratio degrades significantly under the multilayered fabric occlusion of a spacesuit. Castro et al.[5] developed AttenGait, a transformer-inspired model with trainable attention mechanisms, achieving state-of-the-art performance (an average Rank-1 accuracy of 95.8% on the CASIA-B dataset) using optical flow modalities. Despite its rich spatiotemporal feature extraction, its dependence on computationally intensive optical flow and sensitivity to occlusion limit its practicality in resource-constrained extraterrestrial environments. Currently, wearable sensor-based solutions have received widespread attention due to their good environmental adaptability. Such sensors can collect rich gait motion features and enable non-destructive detection. Halim et al.[6] used wearable inertial sensors combined with machine learning algorithms to improve human gait recognition accuracy. However, the support vector machine (SVM) kernel function they used was trained under Earth's gravity

Received date: 2025-01-17
Foundation item: Natural Science Foundation of Shanghai "Science and Technology Innovation Action Plan", China (No. 20ZR1401300)
* Correspondence should be addressed to LÜ Hongzhan, email: lvhz@dhu.edu.cn
Citation: Hao B D, Lü H Z. Human gait recognition based on SOM-K-means clustering for lunar landing spacesuit [J]. *Journal of Donghua University (English Edition)*, 2026, 43(4): 102-109.

constraints. Hsu et al.[7] designed a wearable inertial sensor network and a human motion pattern recognition algorithm using a nonparametric weighted feature extraction and principal component analysis (PCA) dimensionality reduction scheme. Although this reduces computational complexity, some gait feature information is lost during PCA. Sun et al.[8] proposed a laser-engraved wearable gait recognition system using laser-induced graphene pressure sensors embedded in exoskeleton insoles, achieving a recognition accuracy of 99.85% with SVM-based algorithms. While precise in structured environments, its reliance on fixed plantar pressure points may limit adaptability to irregular terrains or dynamic gait variations encountered in lunar exploration scenarios.

Aiming at the specific gait recognition requirements of lunar surface operations, this paper proposes a gait recognition model based on the self-organizing map (SOM)-K-means clustering algorithm. This model is combined with gait data collected in a low-gravity environment to perform human gait recognition. The validity of the model is verified through analysis of the recognition results and comparison with other clustering algorithms.

1 Human Gait Analysis

1.1 Human gait phase classification

First, the gait phases within a single gait cycle during human walking are analyzed. A complete gait cycle commences with the right heel contacting the ground and concludes when the same heel contacts the ground again. This cycle is divided into two phases: the stance (support) phase and the swing phase. In human walking, the stance phase accounts for approximately 60% of the cycle, including pre-support and late support sub-phases, corresponding to heel contact, full-foot landing, and toe-off. The swing phase accounts for about 40%, including pre-swing and late swing sub-phases, corresponding to toe-off, single-leg swing, and foot advancement[9]. From a kinematic perspective in the sagittal plane, during the stance phase, the hip, knee, and ankle joints act synergistically to elongate the lower limb, maximizing the stride length. After initial heel contact, flexion of the knee and ankle cushions the impact. Concurrently, extension of the ankle and knee raises the body's center of gravity to its highest point. In the pre-swing sub-phase, the lower limb shortens and joints flex; in the late swing sub-phase, the lower limb gradually descends and decelerates in preparation for the next heel strike.

1.2 Gait data acquisition and processing

Lunar landing spacesuits operate in an environment significantly different from Earth's surface. Compared to normal gravity conditions on Earth, the low-gravity lunar surface environment causes deviations in movement patterns[10]. Therefore, a simulated low-gravity experimental rig (Fig. 1) was used for gait data acquisition. This rig consists of the following components: a lifting device (low-gravity adjustment device) to assist subjects in achieving a low-gravity state; a supporting frame to ensure the stability of the entire experimental rig; three inertial measurement units (IMUs) to measure and transmit the angular velocity, acceleration, and magnetic field amplitude.

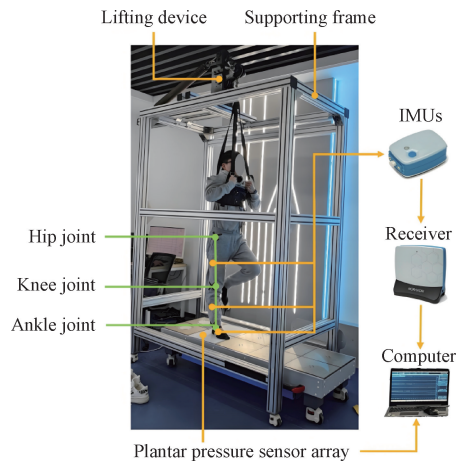


Fig. 1 Experimental rig of gait data acquisition in low-gravity environment

The experiment was meticulously prepared, with rigorous debugging of the low-gravity environment. The IMUs were positioned at the centers of mass of the thighs, calves, and insteps of the subjects' right lower limbs (corresponding to the hip, knee, and ankle joints). The low-gravity adjustment device was then used to gently elevate the subjects in the vertical direction. Concurrently, the plantar pressure was measured in real time using sensors. Once the measured plantar pressure reached one-sixth of the subject's body weight, the subject's height was fixed, thereby successfully creating the desired lunar low-gravity environment. During the experiment, subjects stepped on a platform at a constant speed following a metronome rhythm. The IMUs collected real-time movement data, and a receiver transmitted the data to a computer for real-time viewing, recording, and analysis. Two subjects participated, with heights of 174 and 168 cm and masses of 60 and 55 kg, respectively. The gait data acquisition duration for each subject was approximately 61 s. Considering that astronauts may walk at different speeds during extravehicular operations on the lunar surface, two gait datasets were acquired: slow walking (1.5 km/h) and fast walking (5.0 km/h). Each dataset contains acceleration and angular velocity data for the thigh, calf, and foot segments, resulting in a total of 18 datasets, each containing 24 457 (slow condition) and 24 640 (fast condition) data points.

The angular velocities of the right lower limb segments obtained from the gait data acquisition experiments are shown in Fig. 2 (using slow walking as an example).

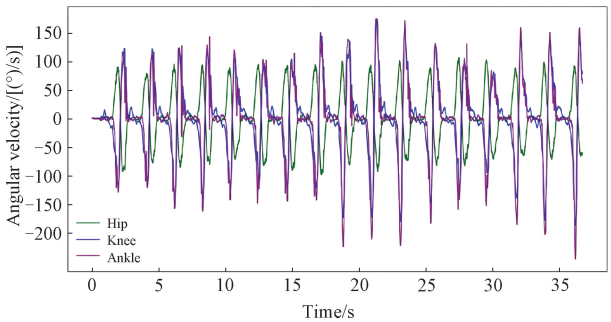


Fig. 2 Angular velocities of the right lower limb segments during gait

It has been established that the gait data vary numerically over time. This is due to friction interference between the sensor and the clothing, which introduces noise into the data collected by the IMUs. Feeding raw data directly into the gait recognition model is detrimental because it reduces both the convergence speed and the overall recognition accuracy. Therefore, data preprocessing prior to model integration is essential.

To address this, data normalization was first applied to the raw data. This process eliminates disparities in magnitude between data variables, ensuring that different variables are given equal weight when influencing the output values. This, in turn, enhances both the convergence speed and the recognition accuracy of the gait recognition model. The Z-score method was selected for data normalization. This technique transforms the data based on the mean and standard deviation, resulting in a distribution with a mean of 0 and a standard deviation of 1. The formula for Z-score normalization is as follows[11].

$$z = \frac{(x - \mu)}{\sigma}, \tag{1}$$

where z is the normalized data; x is the variable in the original data with a mean of μ and a standard deviation of σ.

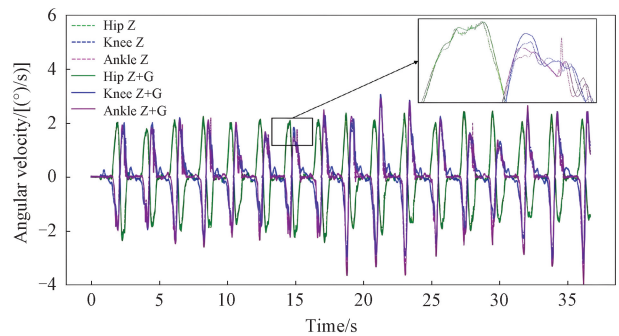
For a feature vector $\boldsymbol{x}$, the normalization is performed component-wise, yielding $\boldsymbol{z} = (\boldsymbol{x} - \boldsymbol{\mu})/\sigma$, where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the corresponding mean and standard deviation vectors.

Subsequently, a Gaussian filter algorithm was applied to the normalized data to extract real and effective signals from the noisy data. This facilitates accurate filtering and noise reduction, thereby reducing noise interference and improving the accuracy of gait features for training the recognition model. Gaussian filtering is a linear filtering method based on the Gaussian function, which determines the weight of each data point in the neighborhood of the central data point. The expression of the Gaussian function is as follows[12].

$$G(u, v) = \frac{1}{2\pi\sigma_G^2}\exp\left(-\frac{u^2 + v^2}{2\sigma_G^2}\right), \tag{2}$$

where u and v denote the offsets from the center data point in the horizontal and vertical directions, respectively; σ_G is the standard deviation of the Gaussian function. In the filtering process, the weight of each data point in the neighborhood is calculated according to the Gaussian function. With each data point as the center, the values of the data points in the neighborhood are multiplied by the corresponding weights and summed to obtain the filtered results.

Figure 3 shows the lower limb gait data after Z-score normalization and Gaussian filtering. As is evident from Fig. 3, the numerical ranges of the original data were converted to the same scale, and the smoothness of the data curves was significantly enhanced. This indicates that Z-score normalization effectively eliminated the differences in scale, and Gaussian filtering effectively eliminated the noise. Thus, these two preprocessing steps significantly enhanced the quality of the gait data, providing a stable and reliable data source for subsequent training of the gait recognition model.



Z— after Z-score normalization; G— after Gaussian filtering processing.

Fig. 3 Angular velocity after pre-processing

2 Gait Recognition Model Based on SOM-K-Means Clustering

The SOM is a specific type of neural network for unsupervised learning. Its purpose is to map high-dimensional input data into a low-dimensional space while maintaining the topology of the data. This is achieved by organizing complex and high-dimensional information (e. g., a dataset containing numerous features) and presenting it in a simple plane or space, thereby making the relationships between the data more intuitive[13]. The SOM can be used to classify and identify data acquired by sensors. The structure of the SOM neural network (Fig. 4) consists of an input layer and a competitive layer (output layer).

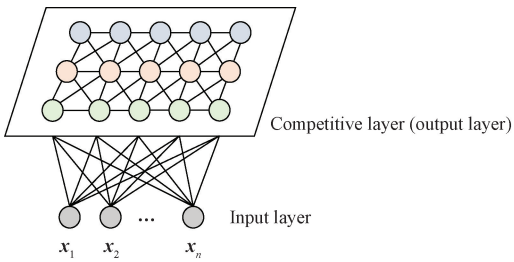


Fig. 4 Structure of SOM neural network

K-means clustering is an unsupervised learning algorithm primarily used to partition samples in a dataset into distinct clusters. The number of clusters must be specified in advance[14]. The fundamental principle is the iterative division of data points into clusters based on proximity to the centroid of each cluster. This iterative process ensures that data points within a cluster exhibit high similarity, while data points from different clusters demonstrate low similarity[14].

The gait recognition model based on SOM-K-means clustering proposed in this paper has the unique advantage of not relying on data labels, as it can classify gaits with similar motion data into the same gait phase. Concurrently, this model effectively compensates for the deficiencies of the standard SOM algorithm, including the difficulty of identifying neurons, the extended training duration, and the over-reliance on the number of clusters and the rationality of the cluster center. The proposed approach thus enhances clustering efficiency while ensuring classification precision. The workflow of SOM-K-means clustering is illustrated in Fig. 5. Initially, gait data are input into the SOM algorithm, initiating the first round of clustering analysis. Subsequently, the elbow rule is implemented to determine the number of clusters k for K-means. The mean value of each category in the SOM clustering results is then designated as the initial cluster center for K-means, and a second round of clustering analysis is conducted to derive the final cluster centers and results.

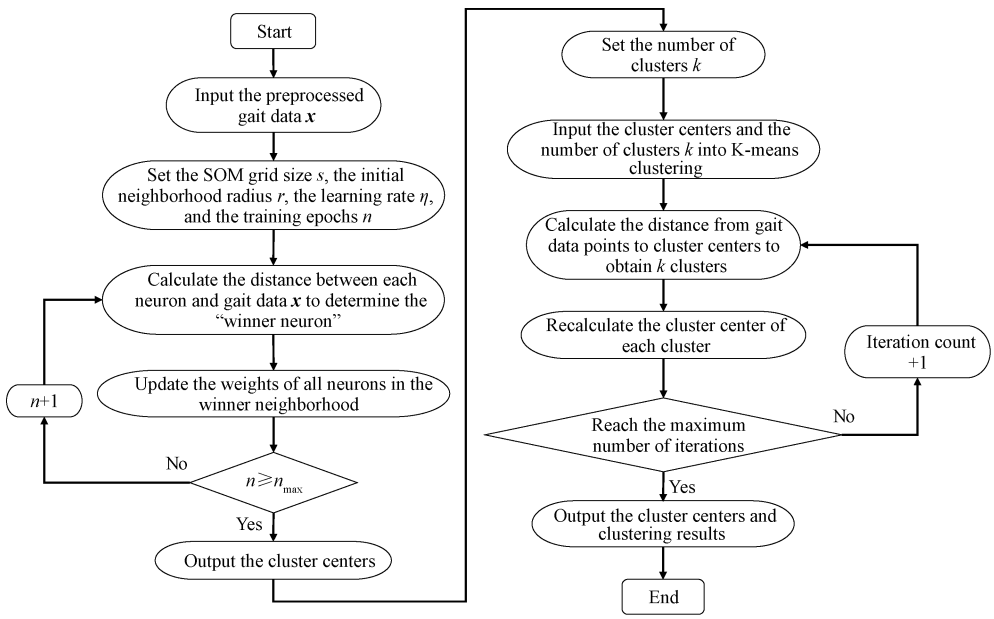


Fig. 5 Workflow of SOM-K-means clustering

The specific steps are as follows.

1) Input the preprocessed gait data $\boldsymbol{x}$.

2) Initialize the SOM neural network parameters, setting the initial weight vector, the grid size s ($s=10\times10$), the initial neighborhood radius r ($r=5$), and the learning rate η ($\eta=0.05$).

3) Output neurons and calculate the Euclidean distance between gait data $\boldsymbol{x}$ and each neuron, and the neuron with the smallest distance is the winning neuron.

4) Determine the winning neighborhood based on the winning neuron and neighborhood range, and update the weights of all neurons in the winning neighborhood.

5) Check whether the maximum number of training epochs is reached; if so, terminate training; otherwise, continue clustering.

6) Set the number of clusters $k=4$, and input the cluster centers and k into the K-means clustering.

7) Calculate the distance between each gait data point and each cluster center, and assign each data point to the cluster whose center is closest. This yields a total of k clusters.

8) Recalculate the cluster center of each cluster.

9) Check whether the maximum number of iterations is reached; if so, terminate the iteration and output the cluster centers and clustering results; otherwise, repeat steps 7) and 8) to continue the iteration.

3 Model Evaluation

3.1 Indicators for model evaluation

The silhouette coefficient and the Davies-Bouldin index (DBI) were used to evaluate the clustering effect of the model.

The silhouette coefficient is a metric that evaluates the clustering quality by quantifying both within-cluster cohesion and between-cluster separation for each data point. Its value range is $[-1,1]$. The closer the value is to 1, the better the clustering effect is and the more accurate the gait recognition is. The calculation formulas are as follows[15].

$$a(i) = \frac{1}{n-1}\sum_{j\in C, j\neq i} d(i,j), \tag{3}$$

$$b(i) = \min\left(\frac{1}{m}\sum_{l\in C'} d(i,l)\right), \tag{4}$$

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}}, \tag{5}$$

$$S = \frac{1}{N}\sum_{i=1}^{N} s(i), \tag{6}$$

where $a(i)$ is the average distance from sample point i to other sample points in the cluster to which it belongs, assuming that sample point i belongs to cluster C; $d(i,j)$ is the distance between i and all other sample points j in the cluster; $b(i)$ is the average distance from sample point i to all other sample points in the other clusters closest to it; cluster C' is the other cluster closest to cluster C; $d(i,l)$ is the distance between i and all sample points l in cluster C'; m is the number of sample points in cluster C'; $s(i)$ is the silhouette coefficient of sample point i; S is the silhouette coefficient of the entire dataset; N is the total number of sample points in the dataset.

The core idea of the DBI is to evaluate the clustering results by measuring the ratio of cluster cohesion and cluster separation. The smaller the value, the better the clustering effect (intraclass cohesion and interclass separation) and the more accurate the gait recognition. The DBI can be calculated by[16]

$$Q_i = \frac{1}{|C_i|}\sum_{x\in C_i} \|\boldsymbol{x}-\boldsymbol{\mu}_i\|_2, \tag{7}$$

$$D_{i,j} = \|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\|_2, \tag{8}$$

$$I_{\mathrm{DB}} = \frac{1}{k}\sum_{i=1}^{k}\max_{j\neq i}\left(\frac{Q_i+Q_j}{D_{i,j}}\right), \tag{9}$$

where I_{DB} is the DBI; Q_i is the intraclass dispersion of the cluster i (the average distance from the samples in the cluster to the center of the cluster); $|C_i|$ is the number of samples in the cluster i; $\boldsymbol{x}$ denotes a sample point in cluster C_i; $\boldsymbol{\mu}_i$ is the center of cluster i (the mean vector); k is the number of clusters; $D_{i,j}$ is the intercluster distance between cluster i and cluster j (the Euclidean distance between the centers of the clusters).

3.2 Model analysis

3.2.1 Model parameter analysis and performance comparison

In the K-means clustering algorithm, the parameter that has the greatest impact on performance is the number of clusters k. The value of k is determined based on the number of gait phases to identify. In the SOM neural network, the grid size s and the initial neighborhood radius r are usually the network parameters that have the greatest impact on performance[17], and the value of s is determined using grid search as follows.

1) Define a range of grid sizes (5×5–20×20).

2) Train the SOM neural network for each size and calculate the quantization error (QE) and topographic error (TE).

3) Select s that minimizes both QE and TE.

After s is determined, a candidate value of r is determined based on a heuristic for grid size: 1/3–1/2 of the diagonal length of the grid, and the same grid search is used to determine the value of r. Finally, $s=10\times10$ and $r=5$ are determined.

To verify that s and r are optimal solutions, the performance of the model with different parameters is compared by using the control variable method. Only the values of s and r are changed, while the input data of the model and other network parameters such as η and k remain unchanged, and the results are shown in Table 1. The clustering performance deteriorates when both s and r decrease, as well as when both increase.

Table 1 Performance of model with different parameters

s	r	Silhouette coefficient	DBI
10×10	3	0.617	0.613
10×10	5	0.881	0.472
10×10	7	0.731	0.552
5×5	5	0.745	0.545
10×10	5	0.881	0.472
20×20	5	0.676	0.582

3.2.2 Model computational complexity analysis

The computational complexity of the SOM-K-means clustering algorithm is mainly affected by the following key parameters: the grid size of SOM, the initial neighborhood size, the number of clusters in the K-means algorithm, and the number of iterations of both algorithms. Improper settings of these parameters can significantly increase the computational complexity, leading to longer computation times and higher costs. In this paper, optimal values for these parameters were determined by a grid search method according to practical needs, resulting in low computational complexity for the

SOM-K-means algorithm. A single run of the algorithm takes only about 90 ms. Given that the average gait cycle duration for an adult is 0.98–1.07 s[18], 90 ms accounts for approximately 9% of a normal gait cycle, indicating that the algorithm is computationally efficient.

3.3 Experimental verification

To visualize the distribution of the gait data, the high-dimensional gait data were projected into three-dimensional (3D) space using PCA for dimensionality reduction. The projection results are shown in Fig. 6.

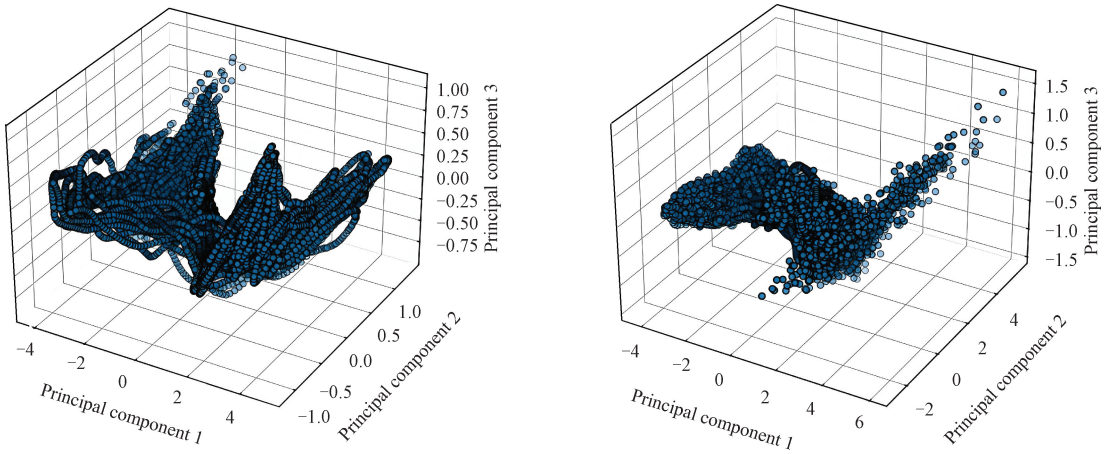


Fig. 6 PCA projection of gait data before clustering: (a) slow walking; (b) fast walking

To verify the validity and generalizability of the proposed gait recognition model based on SOM-K-means clustering, the angular velocity data of the right lower limb segments during slow and fast walking were preprocessed and then input into the model for human gait recognition. The clustering results were further visualized via a 3D PCA projection (Fig. 7). The intra-cluster samples exhibit compactness, and the inter-cluster regions are clearly separated, indicating satisfactory clustering performance on both gait datasets. The silhouette coefficients are 0.881 (slow walking) and 0.873 (fast walking), and the DBIs are 0.472 (slow walking) and 0.476 (fast walking), respectively.

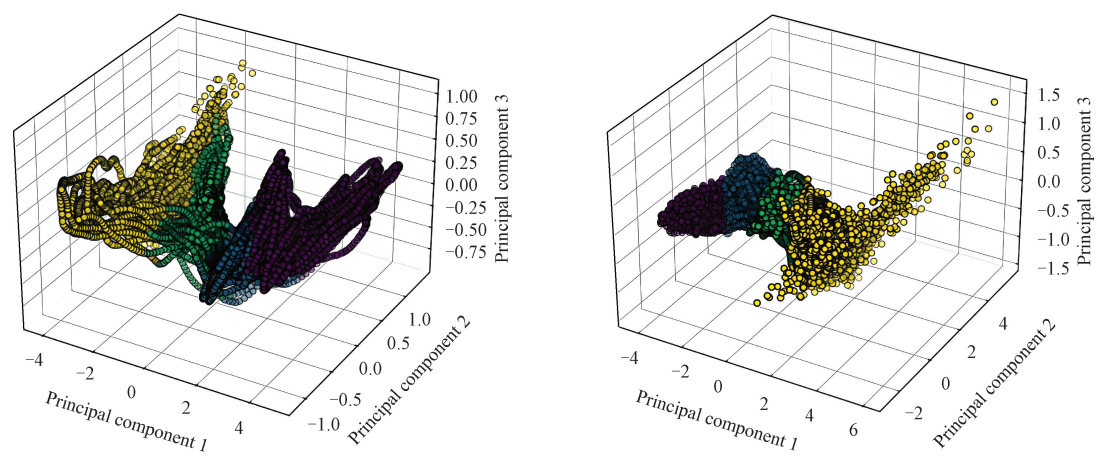


Fig. 7 PCA projection of gait data after SOM-K-means clustering: (a) slow walking; (b) fast walking

The gait recognition results of the SOM-K-means clustering algorithm are shown in Tables 2 and 3. Both gait datasets were divided into four clusters. For the slow walking gait dataset, the data size in Cluster 0 accounts for approximately 20.3% of the entire dataset, Cluster 1 for 31.6%, Cluster 2 for 30.2%, and Cluster 3 for 17.9%. Based on the mean knee angular velocity, Cluster 0 is identified as the pre-swing phase, Cluster 1 as the late support phase, Cluster 2 as the pre-support phase, and Cluster 3 as the late swing phase. For the fast walking dataset, the data size in Cluster 0 accounts for about 18.9% of the whole dataset, Cluster 1 for 28.8%, Cluster 2 for 29.7%, and Cluster 3 for 22.6%. Accordingly, Cluster 0 is recognized as the late swing phase, Cluster 1 as the pre-support phase, Cluster 2 as the late support phase, and Cluster 3 as the pre-swing phase.

According to the average angular velocity of the knee joint, the swing phase has a higher angular velocity than the support phase; the pre-swing phase has a higher angular velocity than the late swing phase; and the pre-support phase has a lower angular velocity than the late support phase. Therefore, the gait recognition results of the SOM-K-means clustering algorithm are reasonable, and the algorithm exhibits good generalization.

Table 2 Slow walking gait recognition results by SOM-K-means clustering algorithm

Cluster number	Data size	Mean knee angular velocity/[(°)/s]	Gait phase
0	4 965	1.90	Pre-swing
1	7 728	0.56	Late support
2	7 386	0.38	Pre-support
3	4 378	1.43	Late swing

Table 3 Fast walking gait recognition results by SOM-K-means clustering algorithm

Cluster number	Data size	Mean knee angular velocity/[(°)/s]	Gait phase
0	4 657	1.84	Late swing
1	7 096	0.59	Pre-support
2	7 318	0.87	Late support
3	5 569	2.42	Pre-swing

Based on the slow walking gait dataset, four more clustering algorithms, namely K-means++, density-based spatial clustering of applications with noise (DBSCAN), deep embedded clustering (DEC) and self-supervised contrastive clustering (SSCC) were used for human gait recognition, and the four clustering algorithms were compared in terms of their silhouette coefficients and DBI. As shown in Table 4, the SOM-K-means clustering algorithm has the largest silhouette coefficient and the smallest DBI.

Table 4 Performance comparison of different clustering algorithms on slow walking gait dataset

Clustering algorithm	Silhouette coefficient	DBI
SOM-K-means	0.881	0.472
K-means++	0.769	0.532
DBSCAN	0.634	0.604
DEC	0.826	0.501
SSCC	0.793	0.519

4 Conclusions

To address the gait recognition problem for astronauts wearing lunar landing spacesuits, a gait recognition model based on the SOM-K-means clustering algorithm was constructed by fusing the SOM and K-means clustering algorithms.

1) Human gait data were collected in a simulated lunar surface low-gravity environment. Z-score normalization and Gaussian filtering methods were applied to preprocess the gait data. This effectively eliminated magnitude differences and noise in the data, providing stable and reliable training data for the gait recognition model, thereby guaranteeing its accuracy and reliability.

2) The SOM-K-means clustering algorithm was used for gait recognition. The effectiveness and generalization of the algorithm were verified by analyzing the recognition results and comparing them with those of other clustering algorithms. The results indicate that the SOM-K-means clustering algorithm for human gait recognition is more accurate.

The gait recognition model based on SOM-K-means clustering proposed in this paper provides a solution to the gait recognition problem in the design of lunar landing spacesuit equipment. Based on the gait recognition model, the lunar landing spacesuit can accurately recognize the astronauts' walking gait and provide the astronaut with a buffer at the moment of lower limb impact.

References

[1] Barclay C D, Cutting J E, Kozlowski L T. Temporal and spatial factors in gait perception that influence gender recognition [J]. *Perception & Psychophysics*, 1978, 23(2): 145-152.

[2] Ryu J, Kamata S I, Ahrary A. SSM-HPC: front view gait recognition using spherical space model with human point clouds [J]. *IEICE Transactions on Information and Systems*, 2012, E95.D(7): 1969-1978.

[3] Han J, Bhanu B. Individual recognition using gait energy image [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(2): 316-322.

[4] Deng M Q, Wang C. Human gait recognition based on deterministic learning and data stream of microsoft kinect [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(12): 3636-3645.

[5] Castro F M, Delgado-Escaño R, Hernández-García R, et al. AttenGait: gait recognition with attention and rich modalities [J]. *Pattern Recognition*, 2024, 148: 110171.

[6] Halim A, Abdellatif A, Awad M I, et al. Prediction of human gait activities using wearable sensors [J]. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2021, 235(6): 676-687.

[7] Hsu Y L, Yang S C, Chang H C, et al. Human daily and sport activity recognition using a wearable inertial sensor network [J]. *IEEE*

Access, 2018, 6: 31715-31728.
[8] Sun M W, Cui S Y, Wang Z Z, et al. A laser-engraved wearable gait recognition sensor system for exoskeleton robots [J]. *Microsystems & Nanoengineering*, 2024, 10: 50.
[9] Perry J, Burnfield J M. Gait analysis: normal and pathological function [M]. 2nd ed. Boca Raton: CRC Press, 2024: 9-16.
[10] Wu X Y, Liu D X, Liu M, et al. Individualized gait pattern generation for sharing lower limb exoskeleton robot [J]. *IEEE Transactions on Automation Science and Engineering*, 2018, 15(4): 1459-1470.
[11] Cheadle C, Vawter M P, Freed W J, et al. Analysis of microarray data using Z score transformation [J]. *The Journal of Molecular Diagnostics*, 2003, 5(2): 73-81.
[12] Ito K, Xiong K. Gaussian filters for nonlinear filtering problems [J]. *IEEE Transactions on Automatic Control*, 2000, 45(5): 910-927.
[13] Salhi M S, Arous N, Ellouze N. Principal temporal extensions of SOM: overview [J]. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2009, 2(3): 121-143.
[14] Bock H H. Origins and extensions of the K-means algorithm in cluster analysis [J]. *Electronic Journal for History of Probability and Statistics*, 2008, 4(2): 1-18.
[15] Bagirov A M, Aliguliyev R M, Sultanova N. Finding compact and well-separated clusters: clustering using silhouette coefficients [J]. *Pattern Recognition*, 2023, 135: 109144.
[16] Ros F, Riad R, Guillaume S. PDBI: a partitioning Davies-Bouldin index for clustering evaluation [J]. *Neurocomputing*, 2023, 528: 178-199.
[17] Wlodarczyk-Sielicka M, Stateczny A. Selection of SOM parameters for the needs of clusterization of data obtained by interferometric methods [C]//2015 16th International Radar Symposium (IRS). Dresden, Germany. Piscataway, NJ: IEEE, 2015: 1129-1134.
[18] Murray M P, Drought A B, Kory R C. Walking patterns of normal men [J]. *The Journal of Bone & Joint Surgery*, 1964, 46(2): 335-360.

面向登月航天服基于 SOM-K-Means 聚类的人体步态识别

郝伯达，吕宏展*
东华大学 机械工程学院，上海 201620

摘　要：登月航天服能否精准识别航天员行走步态，并在下肢受冲击瞬间给予缓冲，成为保障航天员安全作业的关键要素。本文针对此问题展开研究，对人体步态相位进行划分，并在模拟月表低重力环境下采集步态数据。运用 Z-score 标准化和高斯滤波算法进行数据预处理。采用自组织映射（self-organizing map，SOM）-K-means 聚类算法进行步态识别模型的建模，并对模型进行评估。实验结果及聚类算法对比研究验证了 SOM-K-means 聚类算法对步态识别的有效性及模型的泛化性。结果表明，该模型的聚类效果好，步态识别准确，为高性能登月航天服的设计提供了技术支持。

关键词：登月航天服；步态识别；自组织映射；K-means 聚类