

Upper-limb mirror rehabilitation training robot system integrating emotional factors for patients affected by stroke

International Journal of Advanced

Robotic Systems

January–February 2026: 1–14

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17298806261424932

journals.sagepub.com/home/arx



Shijie Liu^{1,*} , Zhen Chen^{2,*} , Lijun Mao¹ , Wenyuan Li¹ , Shi Jie Wang¹ , Ziyang Zhou¹ , Yixue Liu¹ , Yumeng Yang¹ , Yilu Kong¹ , Keshu Cai³ , Jianqing Li¹ , Liang Li¹ and Bin Liu¹

Abstract

Stroke threatens human life and health. After a stroke, patients often experience varying degrees of depression, anxiety, and other emotional symptoms alongside physical hemiplegia. A patient's emotional and psychological states significantly affect their motivation for rehabilitation training. Current robot-assisted stroke rehabilitation focuses more on physical recovery, with little consideration given to the emotional factors affecting patients. This paper proposes an upper-limb mirror rehabilitation training robot system that integrates emotional factors. First, a visual method for motion intention recognition is designed to estimate the end effector pose using visual tags, along with an emotion recognition algorithm based on the Yolo_V11n and ResNet_50 neural network models. Second, emotional factors are incorporated into rehabilitation training to develop a rehabilitation optimization strategy tailored to each patient. Finally, a prototype system is constructed based on a six-degree-of-freedom robotic arm and a binocular vision system. Experimental results show that the proposed system can monitor the subject's facial expressions during rehabilitation training and automatically adjust the training strategy based on their emotional state. The system's average recognition accuracy for the unaffected side's movement trajectory is 4.4 mm, the robot's average tracking accuracy is 4.3 mm, and the emotion recognition model achieves an accuracy of 70% on the validation dataset. The proposed upper-limb mirror rehabilitation training robot integrates the patient's emotional factors into the rehabilitation process, enabling rehabilitation therapists to analyze the patient's psychological activities during training and design personalized physiological and psychological intervention plans.

Keywords

Stroke, robot, mirror rehabilitation, motion intention recognition, emotion recognition

Received: 18 October 2025; accepted: 13 January 2026

Topic: Medical Robotics

Topic Editor: Sunita Bansal

Associate Editor: Yongzhu He

¹The School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

²Department of Clinical Medical Engineering, The First Affiliated Hospital With Nanjing Medical University, Nanjing, Jiangsu, China

³Department of Rehabilitation Medicine, The First Affiliated Hospital With Nanjing Medical University, Nanjing, Jiangsu, China

*Shijie Liu and Zhen Chen contributed equally to this work.

Corresponding authors:

Liang Li, The School of Biomedical Engineering and Informatics, Nanjing

Medical University, Nanjing 211166, China.

Email: liliang@njmu.edu.cn

Bin Liu, The School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China.

Email: liubin@njmu.edu.cn

Data Availability Statement included at the end of the article



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

Stroke is the second leading cause of death and the third leading cause of death and disability combined in the world.¹ By 2050, the number of people dying from stroke globally is expected to increase by 50%.² More than 80% of patients affected by stroke experience upper-limb motor dysfunction,³ which severely affects their daily activities, such as eating, grasping, and object placement. Mirror rehabilitation robots help improve limb function and motor control after strokes.^{4–7} In 2000, Stanford University, United States, proposed an upper-limb mirror rehabilitation training robot system called MIME.⁸ It incorporates the PUMA560 industrial robot, which uses a six-degree-of-freedom position digitizer to detect the movement trajectory of the patient's unaffected upper-limb and then drive the affected side to perform mirror movements within a fixed spatial range at a certain speed. Various upper-limb rehabilitation robot systems integrating mirror therapy have since been rapidly developed. Cui et al. designed a method for identifying upper-limb action intentions based on a combination of inertial motion units (IMUs) and machine learning algorithms.⁹ Wang et al. proposed a control framework for robot-assisted motor learning based on an adaptive frequency oscillator.¹⁰ Lukas Tappeiner et al. developed an upper-limb mirror rehabilitation training system consisting of a cable-driven robot and a visual system.¹¹ Florin Covaciu et al. proposed a combined mirror-EMG robot-assisted therapy system for lower limb rehabilitation, which achieves personalized and immersive rehabilitation through the dual-modal interaction of "mirror therapy + EMG-driven training."⁷

However, the main focus of existing upper-limb mirror rehabilitation robots is patients' physical rehabilitation, with psychological aspects receiving inadequate attention. Anxiety and depression are common psychological health issues occurring after strokes,¹² affecting not only patients' daily mindset but also their rehabilitation progress. Stroke rehabilitation should focus not only on the body's recovery but also on the patient's psychological state, aiming for rehabilitation training that integrates emotional factors.

Emotion-optimized rehabilitation strategies for robotic systems remain in the research phase, with several studies exploring the integration of emotion analysis and affective interaction with rehabilitation protocols. For example, Ghazali et al. proposed a model for emotional states based on human electromagnetic signal information and applied it to rehabilitation robots.¹³ They verified the feasibility of integrating an emotion recognition system with a rehabilitation robot platform. Pepper, developed by SoftBank Robotics, is a robot with emotional recognition and human-robot interaction capabilities.¹⁴ Although it was not specifically designed for rehabilitation, it is often used for the cognitive rehabilitation of elderly and chronic disease patients. These preliminary studies and applications demonstrate the feasibility of integrating emotion into motor and cognitive rehabilitation robots. However, research on upper-limb

rehabilitation robots that integrate emotional factors remains limited.

Recognition of body movement intentions and emotional states is a key component in mind-body integration in robot-assisted rehabilitation. Regarding movement intention recognition, Lu et al. proposed a method for predicting motion intention using deep neural networks combined with one or two IMUs,¹⁵ but the IMU position affects model performance, and drift may occur during prolonged use. Zhang et al. proposed a dynamic adaptive neural network algorithm based on the fusion of multiple features from surface electromyography (sEMG) signals, accurately recognizing eight lower limb movements.¹⁶ Chen et al. decoded human motion intentions using sEMG signals and proposed a hierarchical dynamic Bayesian learning network model.¹⁷ Dong et al. proposed a technique based on the classification of motor imagery electroencephalogram (EEG) signals, achieving high-precision control of lower limb rehabilitation robots and a series of motion intention recognition methods.¹⁸ However, the resolution of EEG limits its independent use for complex motor tasks; multi-modal signal fusion is a feasible approach.

Emotional state recognition mainly relies on physiological signals, such as EEG, electrocardiogram (ECG), and sEMG signals, and methods such as machine vision and multimodal signal fusion.¹⁹ A typical example is the emotion recognition model developed by Elif et al., which combines EEG and physiological data using multimodal signal fusion, machine learning, and deep learning for use in robot-assisted rehabilitation systems.²⁰ Sayed Ismail et al. built an emotion recognition model using support vector machines and a decision tree classifier and evaluated its effectiveness in emotion recognition using ECG signals and photoplethysmography (PPG). The accuracy of using ECG to recognize arousal was 68.75%, whereas the accuracy of using PPG to recognize emotional dimension categories reached 37.01%.²¹ Chanel et al. fused EEG, heart rate, blood pressure, and respiration signals using naive Bayes and Fisher discriminant analysis to recognize the arousal dimension of emotions, achieving an accuracy of 50%–72%.²²

Motion intention and emotion recognition methods based on 1D physiological electrical signals, such as IMU, EEG, and ECG, mainly rely on signal features for indirect estimation; however, signal features are susceptible to interference, and the signal acquisition cost is high. Capturing the affected side's movement trajectory and analyzing facial expression recognition through machine vision are more intuitive and have lower acquisition costs. This paper proposes a vision-based recognition algorithm for upper-limb motion intention to address the above limitations, along with an upper-limb mirror rehabilitation training robot system that integrates emotional factors. The patient's emotional changes during mirror rehabilitation training are monitored to determine their willingness to participate in the rehabilitation process, thereby improving

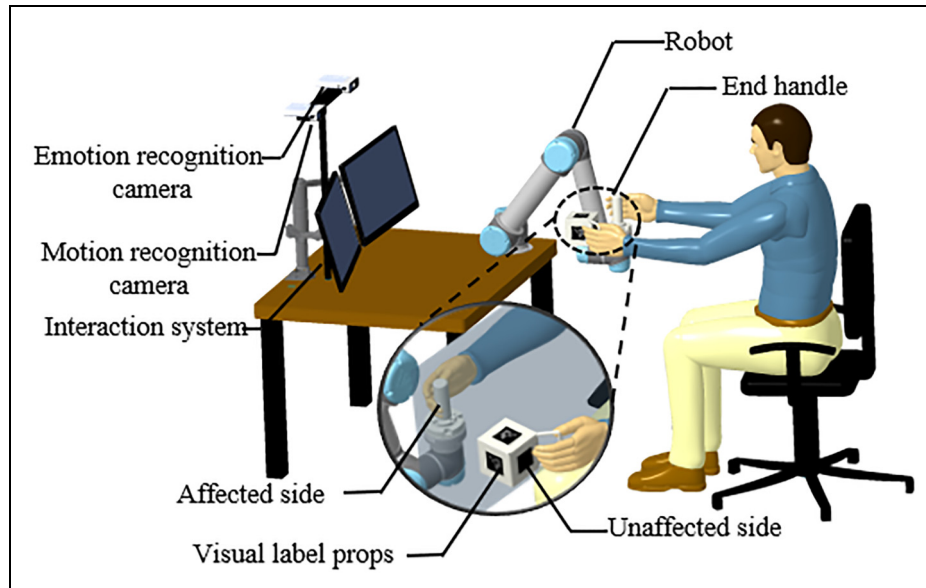


Figure 1. Overall system design diagram.

their rehabilitation training by incorporating their emotional factors.

The rest of this paper is structured as follows: Section 2 describes the hardware architecture design of the system. Section 3 introduces the key technologies of the system, including the motion intention recognition algorithm, emotion recognition algorithm, and rehabilitation optimization strategy. Section 4 validates the system's performance and functionality. Section 5 discusses the implications of the research findings, analyzes the limitations of the current system, and puts forward prospects for future improvement directions. Section 6 summarizes the core conclusions of the entire study, emphasizing the innovation points and practical value of the system.

System overview

This system has two key functions: recognition of motion intention and real time emotion monitoring. First, the patient's motion intention on the unaffected side is recognized accurately. Using the mirror therapy principle, the robot system precisely replicates the unaffected limb's motion trajectory on the affected side, thereby guiding the affected side in rehabilitation training. Second, the patient's emotional state is monitored in real time. According to the patient's emotions, different rehabilitation tasks can be selected to optimize the rehabilitation strategy.

The system structure is shown in Figure 1. It mainly consists of a six-axis robotic arm, two depth cameras, and a workstation. The six-axis robotic arm is the actuator, the cameras are used for motion intention recognition and emotional state monitoring, and the workstation controls the robotic arm and analyzes the emotional states. The workstation is also used to upload the collected training information to a cloud server, allowing rehabilitation therapists to view

the patient's emotional state changes and rehabilitation training data remotely.

The overall working principle of the system is mainly divided into three parts: the recognition, control, and execution systems (Figure 2). The recognition system acquires the patient's motion intention on the unaffected side and a video stream of their facial expression during rehabilitation. The control system analyzes the affected side's motion intention and the facial emotions in the video stream, adjusts the training tasks, and sends control commands to the execution system. Finally, the execution system drives the patient's affected side to perform mirror rehabilitation training.

Method

Motion intention recognition method

In mirror rehabilitation training, the motion intention of the patient's unaffected side is utilized as the reference for training the affected side. Here, a vision system is used to monitor the patient's upper-limb motion intention. As shown in Figure 1, the system captures the motion trajectory of the patient's unaffected side in real time as the patient holds a customized tool with visual tags in 3D space, and estimates the end effector pose through image processing.^{23,24} The occlusion problem caused by the rotation of the visual tag during movement is solved using a 5 cm × 5 cm × 5 cm cubic tool with corresponding visual tags on its surfaces. As long as one side of the cube is not occluded during use, our recognition algorithm can calculate the coordinates of the four corner points of each tag in the pixel coordinate system by determining their center values, following the corresponding OpenCV modules for 3D pose estimation.²⁵ Furthermore, our rehabilitation training tools can be designed according to different

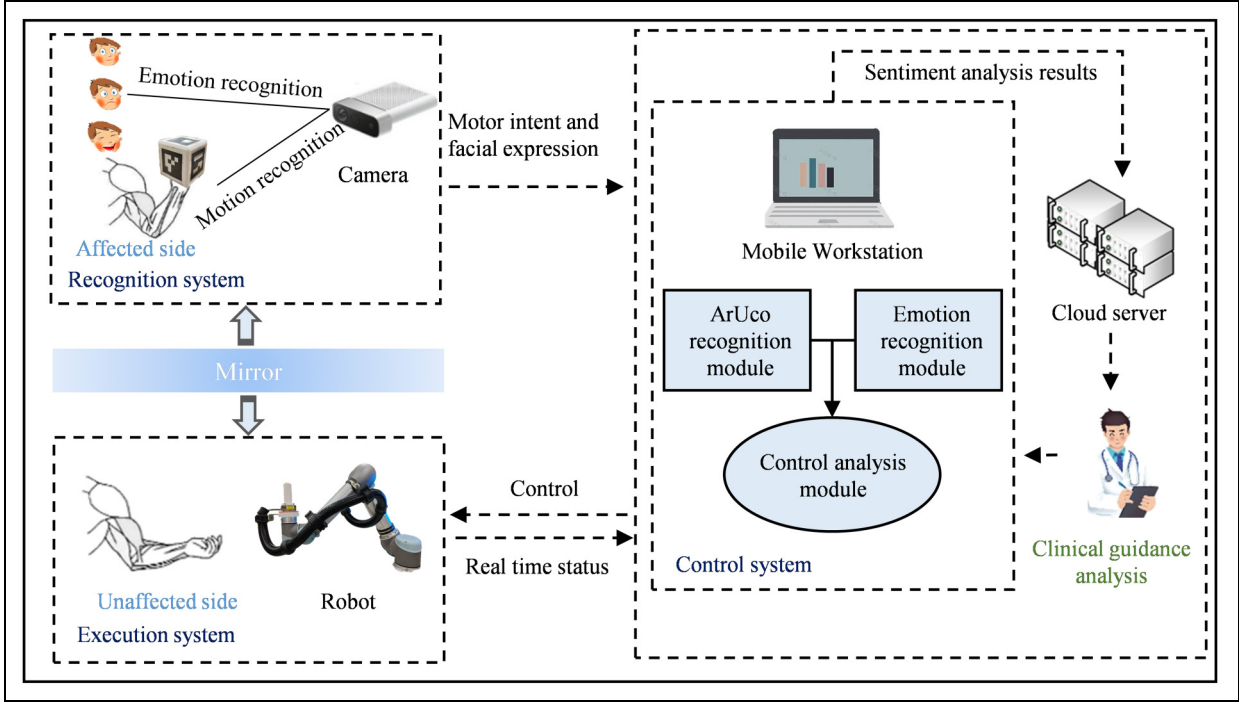


Figure 2. Conceptual diagram of the upper extremity mirror rehabilitation training.

rehabilitation scenarios. As long as the visual tags are attached to the tools, the motion intention of the patient's unaffected side can be detected in real time.

Construction of coordinate system relationship for mirror rehabilitation robot

The operation of the robotic arm is closely related to the coordinate system, which determines the position, posture, and trajectory of each joint and the end effector of the robotic arm. This system relies on the coordinate relationships shown in Figure 3, which mainly include the camera, visual tag, robotic arm base, and robotic arm end effector coordinate systems. The pose of the visual tag in the camera coordinate system is represented by the matrix ${}^{tag}T_{camera}$; the pose of the robotic arm end effector in the robotic arm base coordinate system, ${}^{end}T_{base}$; the transformation relationship between the camera coordinate system and the robotic arm base coordinate system, ${}^{camera}T_{base}$; the transformation relationship between the visual tag and the robotic arm end effector, ${}^{tag}T_{end}$. This system infers the patient's movement intention by recognizing the motion trajectory of the handheld visual tag and then uses its inference for the mirror control of the robotic arm (Section 3.1). To implement this control function, the system should accurately obtain the spatial coordinate transformation relationship between the visual tag and the robotic arm base. With the pose of the handheld visual tag in the camera coordinate system denoted as ${}^{camera}T_{tag}$ and the pose of the visual tag in the robotic arm base coordinate

system ${}^{base}T_{tag}$ obtained using (1), where ${}^{base}T_{camera}$ is the system's inherent hand-eye transformation matrix (which can be obtained through hand-eye calibration), ${}^{base}T_{end}$ and ${}^{camera}T_{tag}$ is directly obtained through the vision system and robotic arm control system.

$${}^{base}T_{tag} = {}^{base}T_{camera} \bullet {}^{camera}T_{tag} \quad (1)$$

In this study, ${}^{base}T_{camera}$ is obtained through hand-eye calibration, and its specific implementation is shown in

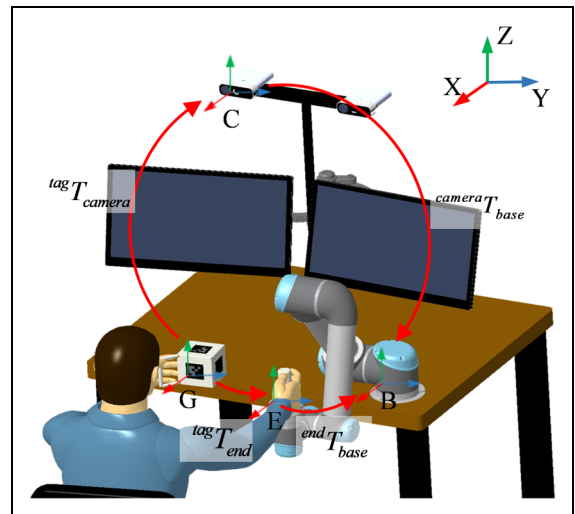


Figure 3. System coordinate transfer diagram.

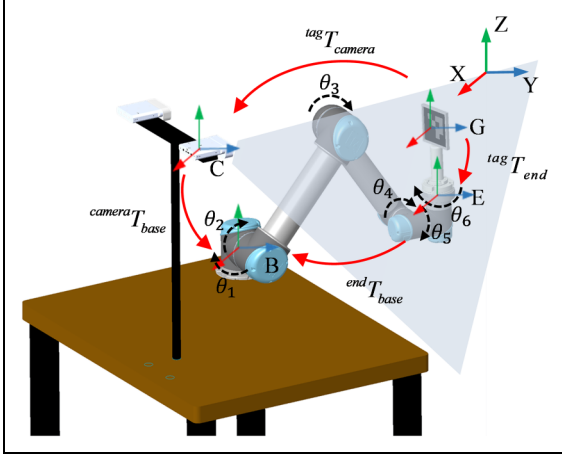


Figure 4. Diagram of hand-eye calibration coordinate relationship.

Figure 4. During calibration, the visual tag is rigidly fixed at the center position of the robotic arm end effector flange, and the robotic arm is controlled to perform random sampling movements within the 3D workspace. The system synchronously collects the corresponding ${}^{base}T_{end}$ and ${}^{camera}T_{tag}$ at each sampling point and constructs a calibration dataset for subsequent hand-eye relationship calculation.

Given ${}^{base}T_{camera}$ as the system's inherent hand-eye relationship matrix, (2) and (3) are derived. The solution of (3) is equivalent to that of a typical $AX=XB$ problem. In this study, the least squares method of Tsai²⁶⁻²⁸ is used to solve for ${}^{end}T_{tag}$, and ${}^{base}T_{camera}$ is obtained from (2)

$${}^{base}T_{camera} = {}^{base}T_{end} \bullet {}^{end}T_{tag} \bullet ({}^{camera}T_{tag})^{-1} \quad (2)$$

$$\begin{aligned} & {}^{base}T_{end_i} \bullet {}^{end}T_{tag} \bullet ({}^{camera}T_{tag_i})^{-1} \\ &= {}^{base}T_{end_j} \bullet {}^{end}T_{tag} \bullet ({}^{camera}T_{tag_j})^{-1} \end{aligned} \quad (3)$$

where i and j represent different sampling points ; $({}^{base}T_{end_i})^{-1} \bullet {}^{base}T_{end_i}$ is A, and $({}^{camera}T_{tag_i})^{-1} \bullet {}^{camera}T_{tag_i}$ is B.

Mirror trajectory following control

According to the above results, the system obtains the transformation matrix ${}^{base}T_{camera}$ from the visual camera coordinate system to the robotic arm base coordinate system. The spatial position and pose of the patient's unaffected side, which holds the visual tag in the robotic arm base coordinate system, are derived using (1). The system's mirror plane is set as the human body vector plane, and the rehabilitation training scenario is preconfigured as follows: the patient sits at the center of a fixed seat directly in front of the rehabilitation training platform, with the robotic arm assisting in mirror training. Since the core objective of this training task is to achieve

motion mirroring between the affected and unaffected limbs, rather than strictly adhering to anatomical mirroring principles, the center of the training platform is defined as the human body vector plane once the patient is seated on the fixed seat. Therefore, the patient's trunk movement will not affect the mirroring function of the system.

After the mirror transformation is performed on the spatial pose of the unaffected side (holding the tag), it serves as the target pose ${}^{base}T_{affect}$ for the affected side. The motion control of the robotic arm then achieves the real time following and reproduction of the unaffected side's movement trajectory for the affected limb.

The mirrored position of the unaffected side (holding the tag in the robotic arm base) is calculated using (4).

$${}^{base}T_{affect} = {}^{base}T \bullet {}^{unaffected}T_{affect} \quad (4)$$

where ${}^{unaffected}T_{affect}$ is the mirror transformation matrix.

The robotic arm's mirror-following control is implemented as follows:

The pose of the unaffected side (holding the tag in the robotic arm base coordinate system) ${}^{base}T_{unaffected}$ is obtained based on the motion intention recognition system. The system also obtains the mirrored pose of the unaffected side using (4), which represents the real time pose of the affected side, which is driven by the robotic arm in the robotic arm base coordinate system ${}^{base}T_{affect}$. These are the target position ${}^{base}T_{target}$ and current position ${}^{base}T_{current}$ in the control process, respectively, and their discrepancy is calculated. The components of ${}^{base}T_{target}$ and ${}^{base}T_{current}$ are

$${}^{base}T_{current} = \begin{pmatrix} R_{current}^{base} & t_{current}^{base} \\ 0, 0, 0 & 1 \end{pmatrix} \quad \text{and} \quad {}^{base}T_{target} = \begin{pmatrix} R_{target}^{base} & t_{target}^{base} \\ 0, 0, 0 & 1 \end{pmatrix}, \text{ respectively.}$$

The deviation vector δ^{base} of the target position relative to the current robotic arm base coordinate system is calculated using (5).

$$\begin{aligned} \delta^{base} &= \Delta({}^{base}T_{target}, {}^{base}T_{current}) \\ &= \begin{pmatrix} t_{target}^{base} - t_{current}^{base} \\ \text{vex}(R_{target}^{base} R_{current}^{base} - I_{3 \times 3}) \end{pmatrix} \end{aligned} \quad (5)$$

where δ^{base} is a 6×1 vector where the first three rows are position offsets and the last three rows are angular deviations.

The incremental control step size is denoted as λ , which is then used to compute the next control target increment.

$$\Delta\delta^{base} = \lambda \bullet \delta^{base} \quad (6)$$

Once the next control target increment is known, the robotic arm's target position is updated using (7).

$${}^{base}T_{update} = \begin{pmatrix} (I_{3 \times 3} + s(\Delta\delta^{base}(4:6))) & t_{current}^{base} + \Delta\delta^{base}(1:3) \\ 0, 0, 0 & 1 \end{pmatrix} \quad (7)$$

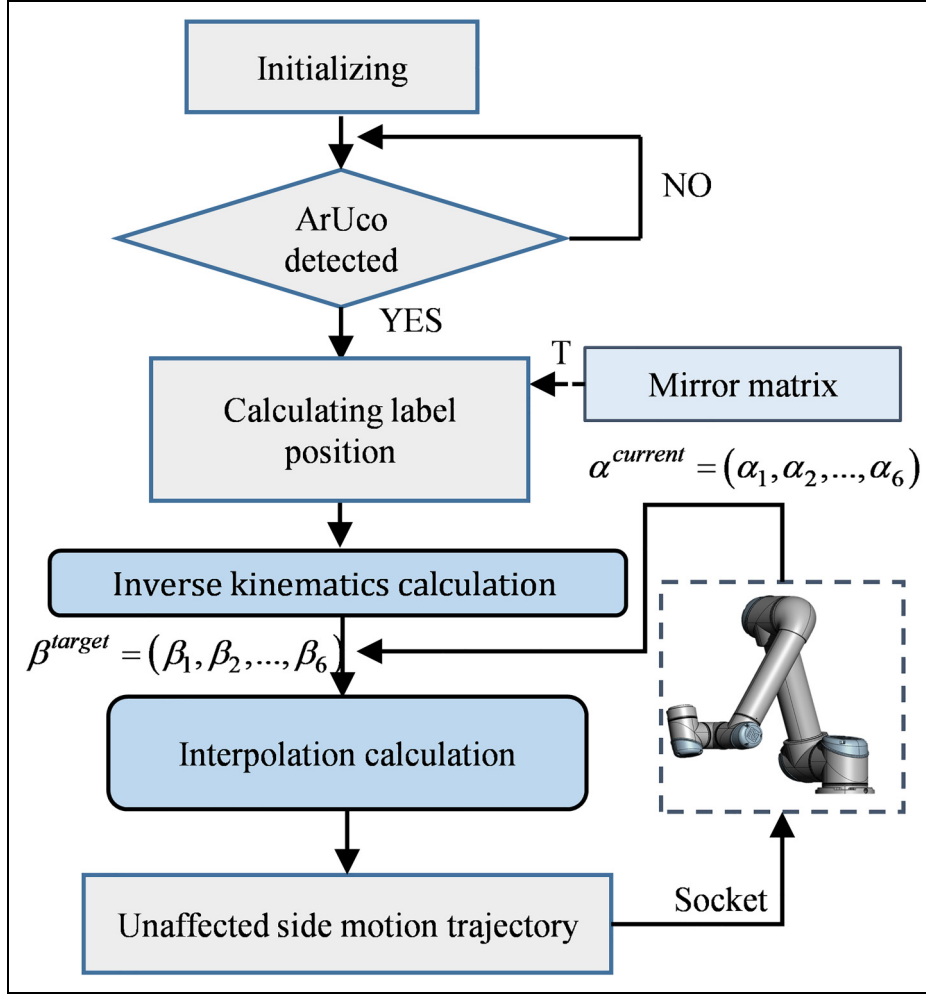


Figure 5. Mirror rehabilitation control framework.

Interpolation planning is performed between ${}^{current}T_{base}$ and ${}^{update}T_{base}$ to ensure the smooth motion of the robotic arm as it approaches the target position. As shown in (8) and (9), a smooth trajectory C_t is obtained between each segment of the deviation, thereby achieving smooth motion.

$$\alpha^{current} = IK({}^{base}T_{current}) = (\alpha_1, \alpha_2, \dots, \alpha_6) \quad (8)$$

$$\beta^{target} = IK({}^{base}T_{target}) = (\beta_1, \beta_2, \dots, \beta_6) \quad (9)$$

where IK refers to an inverse kinematics algorithm,^{29,30} $\alpha^{current}$ represents the joint angles of the current posture, and β^{target} represents the joint angles of the target posture.

$$R_k = (1 - r_k)\alpha^{current} + r_k\beta^{target}, r_k = \frac{k-1}{m-1}, k = 1, 2, \dots, m \quad (10)$$

where R_k represents each interpolation point between trajectories, m is the number of interpolation steps, k is the number of trajectory points, and r_k is the interpolation parameter.

Finally, the system dynamically updates the control variable C_t and transmits it to the robotic arm controller via the Socket communication protocol. It then drives the robotic arm end effector to approach the target position gradually, thereby achieving the mirrored reproduction of the unaffected side's movement trajectory for the affected limb. Figure 5 shows the system control flow.

Patient emotional state recognition algorithm

This system combines the Yolo_V11n (nano version) network model^{31,32} with the ResNet_50 network model³³ to recognize the patient's emotional state during rehabilitation based on their facial expressions. We adopted the official implementation of Yolo_V11n released by Ultralytics, with the specific version specified as Ultralytics Yolo_V11n v8.3.0. Yolo_V11n is commonly used for object detection but can be applied to face detection tasks. It not only recognizes faces in images but also estimates poses, such as facial orientation, making it highly suitable for emotional state recognition in

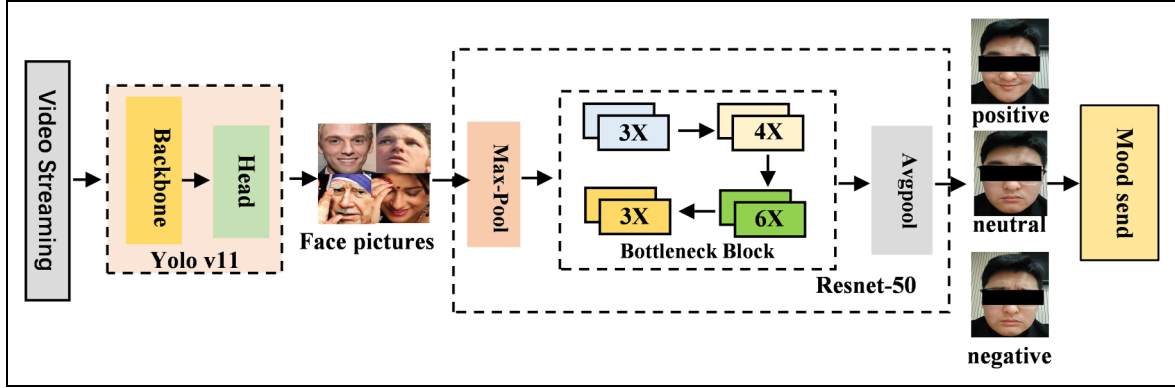


Figure 6. Architecture of emotion recognition program.

this study. In this research, we convert the bounding box annotations from the WIDER_FACE dataset³⁴ from absolute pixel coordinates to the relative coordinate format required by the Yolo_V11n model. Specifically, we compute and transform each bounding box's center coordinates and dimensions into proportions relative to the image's width and height. The converted label files adhere to the Yolo_V11n format and are saved alongside the image files, providing a standardized dataset for the subsequent object detection task. ResNet_50 is widely used for image classification. It excellently recognizes object categories in images and can be used for facial feature extraction, capturing subtle differences in facial images. Therefore, it is appropriate for predicting emotional states based on facial expressions in this study.

The emotion recognition process is shown in Figure 6. First, the video stream captured by the camera is input into the program, where the trained Yolo_V11n model detects and labels the face in the video. Then, the labeled face image is fed into the trained ResNet_50 model for facial expression state analysis. Finally, the system classifies the patient's facial expressions as indicating positive, neutral, or negative emotions and labels the analysis results in the video stream in real time.

Rehabilitation optimization strategy integrating emotional factors

The patient's emotions influence their level of participation and effort in rehabilitation tasks, thus affecting the effectiveness and progress of rehabilitation. A patient who is in a positive, optimistic state will be more willing to participate in rehabilitation training and less likely to develop resistance. However, a patient in a negative state will exhibit low cooperation in rehabilitation tasks and be more likely to develop resistance. Therefore, this study focuses on ensuring patient safety during rehabilitation, and an initial attempt is made at integrating emotional factors into rehabilitation optimization strategies. However, extensive clinical exploration and practice are required for

the rational, effective integration of emotions into rehabilitation optimization.

Three emotional states, namely, positive, neutral, and negative, are used as conditions in this study, with each emotion corresponding to a unique adjustment strategy. If the patient is in a positive emotional state, then the system will appropriately increase the training intensity or duration. If the patient is in a neutral state, then the system will continue the established training plan to ensure rehabilitation effectiveness. If the system detects negative emotions lasting more than 5 s or accounting for over 80% of the emotional state within 10 s, it will immediately stop the training and issue a warning message. Figure 7 illustrates the control logic through which the system exerts a key role in rehabilitation optimization by incorporating emotional factors, thereby providing patients with a safer, more efficient, and personalized rehabilitation experience.

In addition, we have implemented two supplementary safety measures: First, an emergency stop button is installed. In the event of a potential hazard during training, the therapist can press the emergency stop button promptly to shut down the system immediately. Furthermore, the end-effector handle is designed to be detachable at any time, allowing the patient to disengage from the system quickly. Second, a safety force threshold is preset in the system, with an initial value of 40 N, which can be personalized adjusted according to individual patients' conditions. If the patient's resistance force exceeds this threshold, the system will trigger an emergency shutdown. Moreover, throughout the entire training process, a therapist is assigned to provide full-time supervision to prevent hazardous situations. These measures effectively prevent potential risks caused by emotional fluctuations.

Result

Prototype construction

As shown in Figure 8, the system uses the six-axis robotic arm UR5 (Universal Robots) as the actuator. The depth-

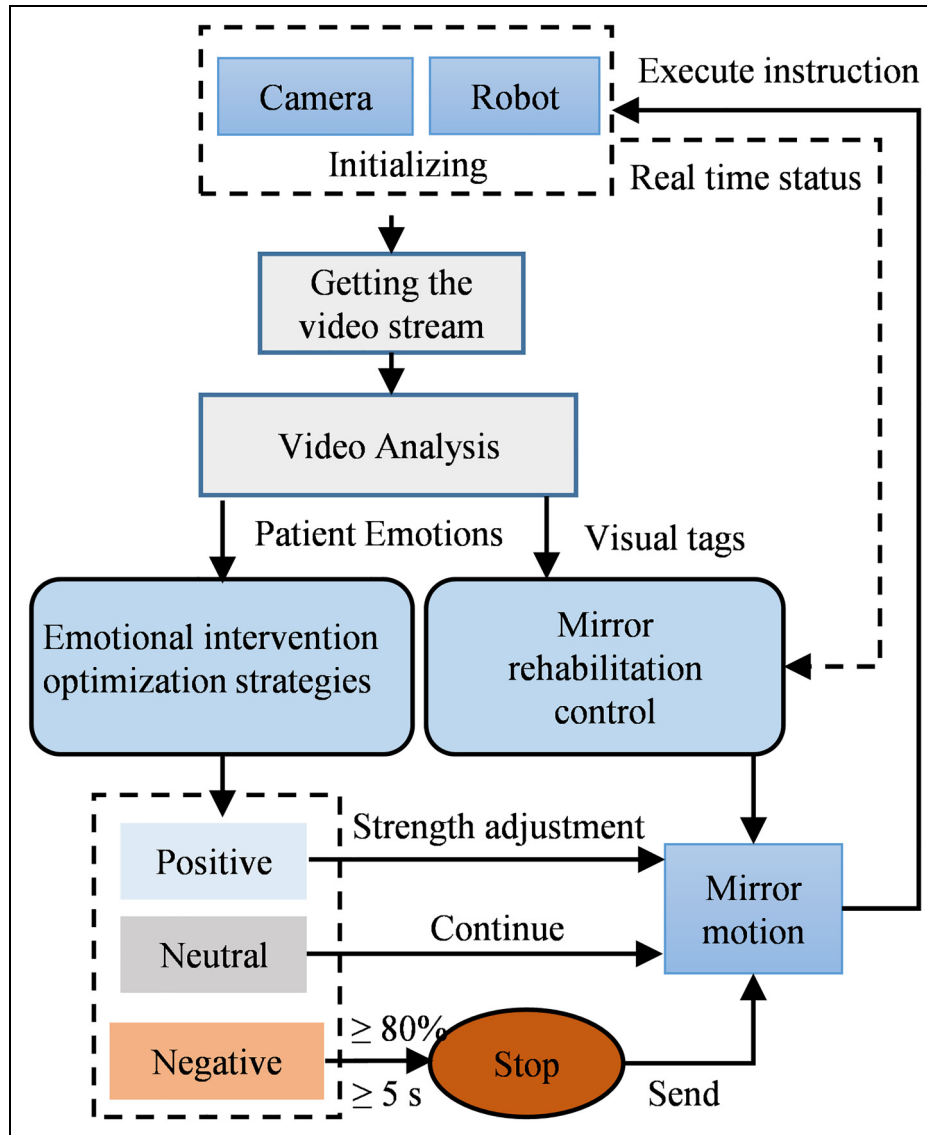


Figure 7. Logic diagram of emotional control strategies.

sensing sensor Azure Kinect DK (Microsoft) is used as the video frame acquisition device for the motion intention and emotion recognition programs. The video stream is transmitted to a mobile workstation, which communicates with the actuator through a router. The mobile workstation also uploads patient training data and emotional state information during rehabilitation to the server, allowing rehabilitation therapists to access various data, such as emotional changes and rehabilitation progress.

Performance validation

Motion intention recognition accuracy test. The performance of our visual tag recognition system is assessed accurately by installing the visual tag on the end effector of the robotic arm, recording 120 data points of the visual tag's position in

a stationary state, and analyzing them. As shown in Figure 9, the system is set to perform periodic repetitive motion along a rectangular closed loop path with a preset width of 200 mm and a length of 250 mm. Using the depth camera as the recognition tool and our motion intention recognition method, we precisely calculate and record the coordinates of the center of four fixed key points' tags in the camera coordinate system per motion cycle. This process covers 35 complete motion cycles, with a total of 140 key points of data collected. Subsequently, we calculate the distance between the lateral and longitudinal directions per cycle.

Within the preset experimental range, the recognition program's average error of position recognition in the static state is 1.5 mm. During motion, the average error of motion recognition per path segment is 4.4 mm (Table 1).

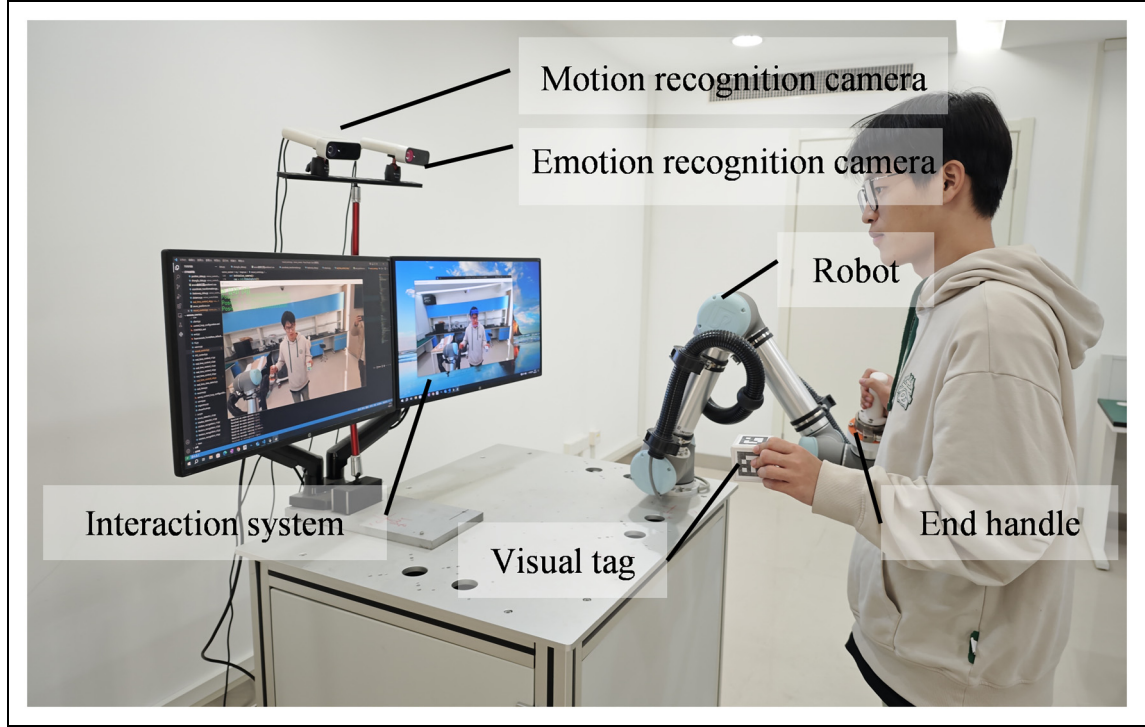


Figure 8. Prototype of overall system.

Therefore, the accuracy of visual tag recognition meets the technical requirements of this study.

Emotion recognition validity test. For the training of the emotion recognition model in this study, the public WIDER FACE dataset³⁴ and preprocessed AffectNet dataset³⁵ were adopted as the training data for face detection and emotion recognition tasks, respectively. For the Yolo_V11n model, the training process was configured with 50 epochs and a batch size of 32, and the SGD optimizer was employed to enhance convergence efficiency, utilizing an initial learning

rate of 0.01 and a momentum of 0.9. Prior to the training of the ResNet_50 model, preprocessing operations on the dataset had been completed in advance. Specifically, the eight-class facial images in the AffectNet dataset are reclassified: happiness and surprise as positive emotions; sadness, anger, disgust, fear, and contempt as negative emotions; and neutral as a neutral emotion. Each category contains 12,000 images, which serve as the training data for the model. The ResNet_50 model was trained for 40 epochs using a batch size of 8. The training employed the SGD optimizer with an initial learning rate of 0.001 and a momentum of 0.9. The training results show an accuracy of 70%; Figure 10 shows the confusion matrix of the model and the multiclass receiver operating characteristic (ROC) and area under the ROC curve.

To validate the effectiveness of the emotion recognition algorithm in this study, we edit movie clips with three different emotional tones (calmness, sadness, and happiness) from the films *Green Book*, *Manslaughter*, and *Home Alone*, each approximately 6 min long. Wearing headphones, three participants in a calm state watch the three movie clips. Meanwhile, our emotion recognition system analyzed and recorded their

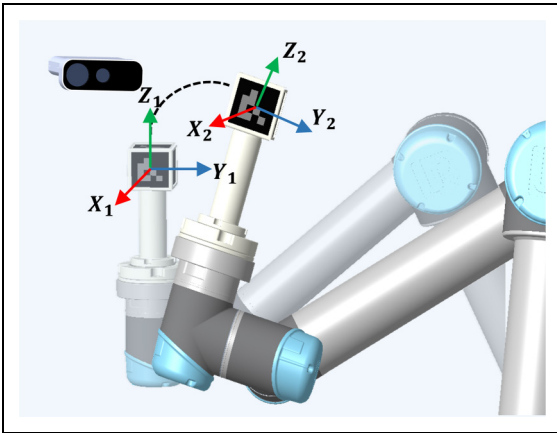


Figure 9. Motion intention recognition accuracy test.

Table 1. Analysis of statistical indicators.

Statistical indicators/route direction	Stationary (mm)	Vertically (mm)
AV	1.5	4.4
SD	1.3	2.0

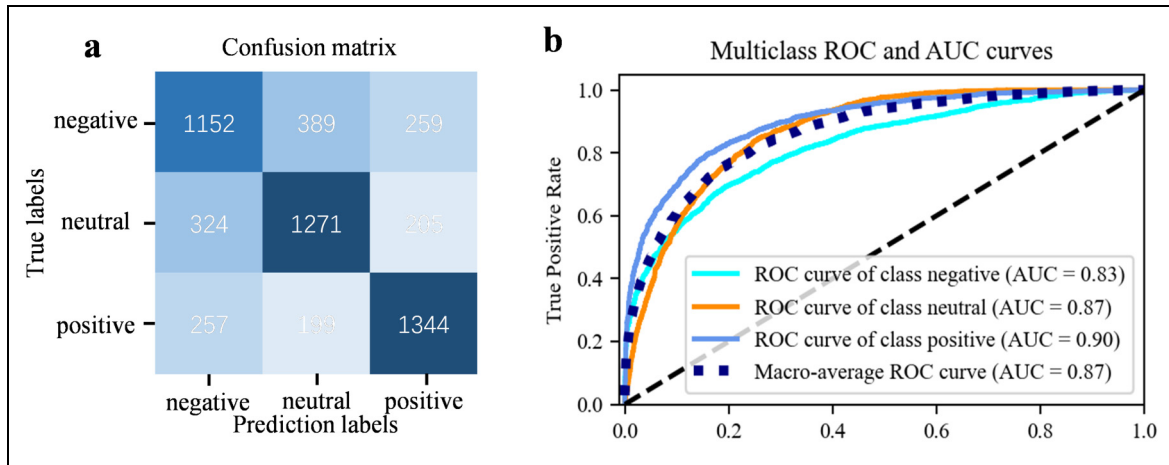


Figure 10. (a) Confusion matrix of the ResNet50-based emotion recognition model, and (b) the multiclass ROC curves and corresponding AUC values.

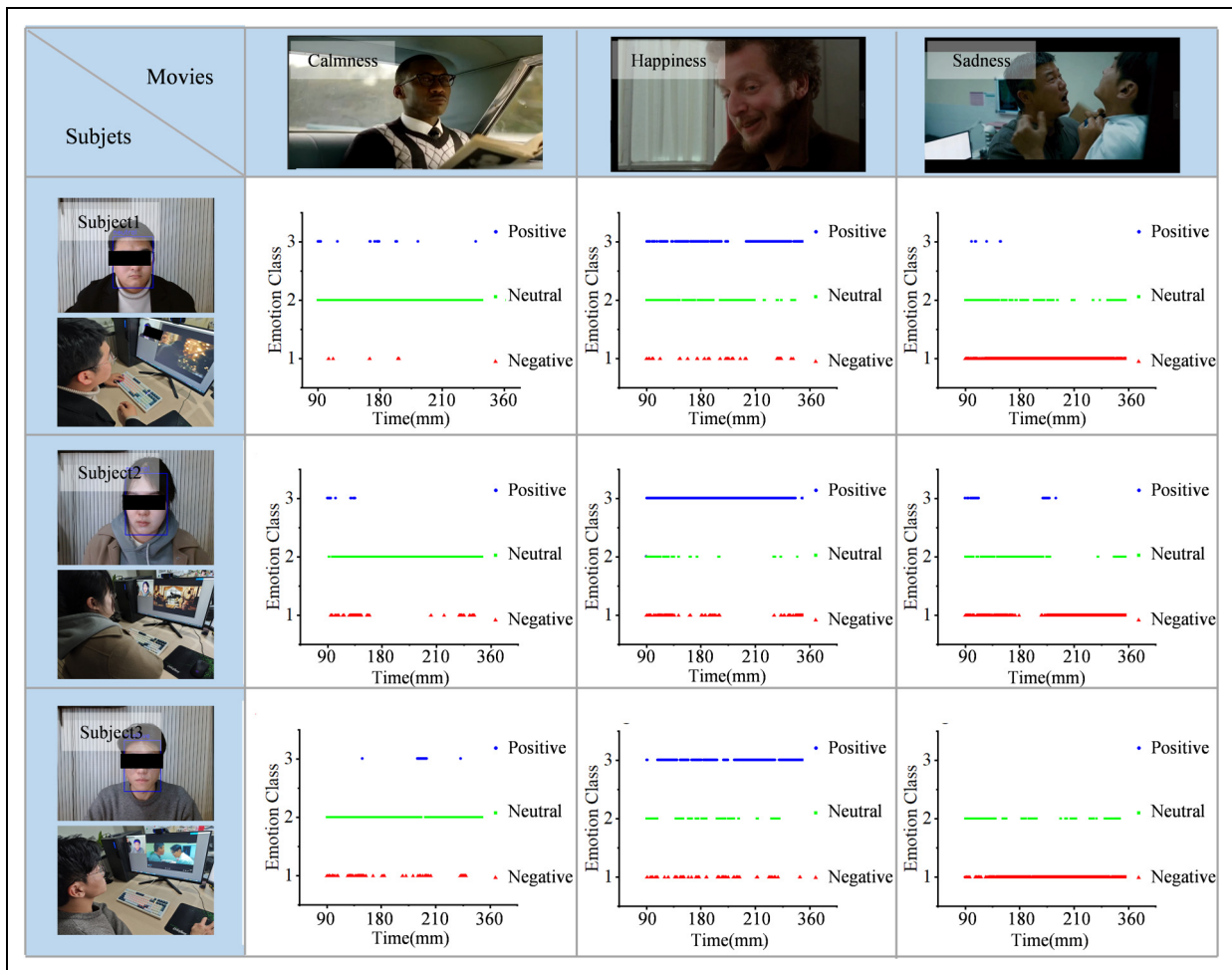


Figure 11. Emotion recognition feasibility analysis for watching different movie clips. The rows show the emotional changes of three subjects watching three different emotional colors.

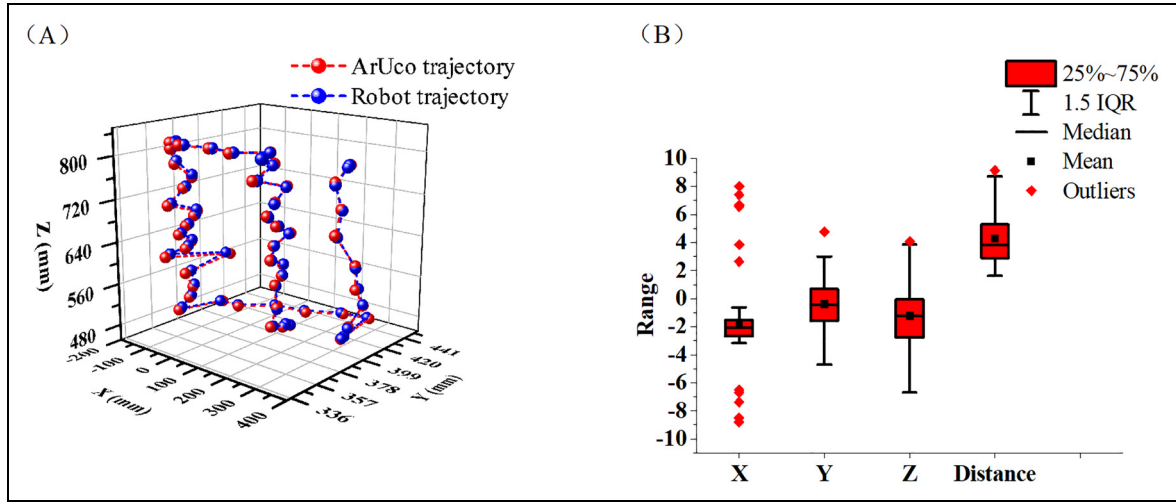


Figure 12. (a) Comparison graph between prop motion trajectory and tracking trajectory of end of robotic arm after origin alignment. (b) Result of statistical analysis of tracking error.

emotional states at a sampling frequency of 8.33 Hz during the viewing process, and then compared the emotional fluctuations of the three participants to validate the program's effectiveness. The results (Figure 11) indicate that the emotional fluctuations of the participants while watching the three movie clips match the emotional tones of the clips, with the algorithm performing excellently in analyzing negative emotions.

Robotic arm trajectory mirror tracking accuracy test. To test the trajectory tracking accuracy of the robotic arm, in this experiment, we place the visual tag tool on a stand and move it in space, with the robotic arm following the movement in real time. The system then recognizes and records the motion trajectory of the visual tag at a sampling rate of 60 Hz. Then, the mirrored trajectory of the tag is obtained through hand-eye calibration and the mirror transformation matrix, with the time. A comparison of the trajectory of the visual tag and the robotic arm's tracking trajectory error, along with the statistical analysis results, is displayed in Figure 12.

In the selected area of this experiment, the average distance error of the robotic arm's trajectory tracking is 4.3 mm, with a standard deviation of 1.9 mm (Table 2). Therefore, the robotic arm's tracking precision in this system roughly meets the accuracy requirements for the rehabilitation training outlined in this study.

Table 2. Statistical indicators of X, Y, and Z axis tracking errors for props and robotic arm ends.

Statistical indicators/axis	X (mm)	Y (mm)	Z (mm)	Distance (mm)
AV	-1.8	-0.3	-1.2	4.3
SD	3.0	1.7	2.3	1.9

Functional validation

In order to verify whether the system meets the expected standards, we invited three adults as subjects to conduct functional verification of the system. During the trial, the subjects' limb movements and emotional states during training were recorded via cameras. Results within a 10-s window were extracted and presented. In this process, to test the feasibility of emotional state recognition during system operation, we required the subjects to intentionally exhibit three types of emotions (negative, neutral, and positive) in sequence. Before the experiment began, the participants were informed of the specific functions of this system and signed informed consent documents. This study has passed the review of the Ethics Committee of Nanjing Qixia District Hospital (Report No. 20220609), and the Ethics Committee approval date is June 9, 2022. Figure 13 shows the system assisting patients with rehabilitation training while its emotion recognition module analyzes, displays, and records the patients' real-time emotional states. Findings suggest that this system can aid in robot-assisted mirror rehabilitation training while monitoring the subject's facial expressions and implementing emotion-based intervention during training, ensuring patient safety throughout the activity. This result indicates the initial success of our approach in integrating emotional factors into upper-limb mirror rehabilitation training.

Discussion

This paper proposes an upper-limb mirror rehabilitation training robotic system integrated with emotional factors and verifies its performance and feasibility. The system monitors patients' emotional states while providing robot-assisted rehabilitation training support. Specifically,

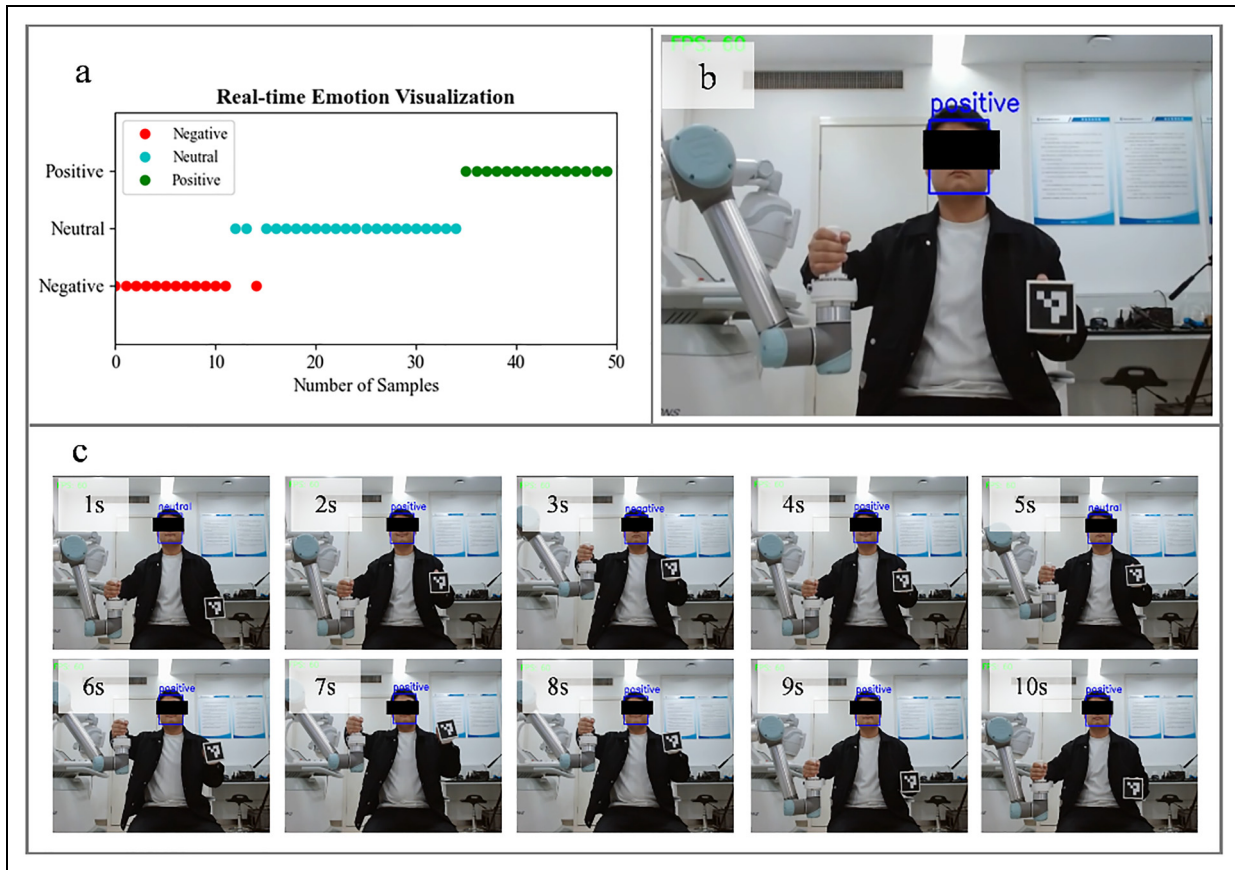


Figure 13. System function verification: (a) Real-time visualization of emotional states during training; (b) backend display of subjects' training videos; and (c) overview of upper-limb rehabilitation tasks performed by subjects in 1–10 s.

machine vision is adopted to track the motion trajectory of the patient's unaffected limb, and facial expression analysis is implemented to assess emotional state, this emotion recognition module consists of two sequential steps: first, Yolo_v11 is used to detect and localize human faces from video streams, which is a well-established and mature technology; second, ResNet_50 is employed to analyze the identified facial regions and determine the corresponding emotional states, effectively avoiding the interference of different environmental factors on the detection results. Through the above mechanisms, the system achieves upper-limb mirror rehabilitation training that incorporates emotional factor regulation.

For the first time, this study attempts to incorporate emotional factors into robot assisted upper-limb mirror rehabilitation by using machine vision to capture motion intentions and emotional states, thereby realizing a closed-loop intervention mode of “motion intention capture – real-time emotional state monitoring – dynamic adaptation of rehabilitation strategies.” This innovative design breaks through the limitation of traditional rehabilitation robots that only focus on motor function recovery. The system can not only accurately respond to the patient's movement instructions, but also adjust training or trigger safety protection mechanisms according to emotional

fluctuations. In terms of safety assurance, the system has established a triple collaborative protection mechanism, which specifically includes an emergency stop button, a personalized safety force threshold, and an emotion-driven emergency shutdown measure, forming a comprehensive safety barrier to ensure the safety of patients during training.

It is worth noting that the determined motion intentions and emotional states can only partially reflect the true thoughts of patients. Physiological signals, such as IMU, EEG, and ECG signals, can reflect a patient's motion intention and emotional state,^{16–18} but relying on a single modality may compromise accuracy and make precise recognition difficult. Therefore, multimodal fusion techniques for motion intention and emotional state recognition should be explored for accurate identification.

Performance and functional verification results demonstrate that the accuracy of motion intention recognition and emotion recognition of the system has met the expected standards, which can satisfy the requirements of routine upper limb rehabilitation training. Notably, integrating emotional factors into the rehabilitation process and implementing intervention optimization based on emotions is a cutting-edge research direction with both complexity and challenges, and the research team is still in the preliminary

exploration stage of this field. Based on this, in the initial phase of the research, we completed the system functional verification by recruiting healthy subjects. Its core goal focused on verifying the feasibility and safety of the system's core technologies, laying a solid foundation for the subsequent clinical transformation of the technologies. Meanwhile, the research group deeply recognizes the critical significance of clinical verification for the implementation of the research, has initiated cooperation with relevant clinical institutions to advance the preparation of clinical trials, and has successfully completed the special tests on safety and performance conducted by third-party institutions. At this stage, we are recruiting clinical subjects among patients with upper limb hemiplegia after stroke, aiming to evaluate the practical application effectiveness of the system in real clinical scenarios.

In future research, we will further integrate multimodal physiological data to improve the accuracy of emotion analysis; simultaneously supplement and improve clinical trial data to continuously enhance the clinical value and transformation potential of the research results. In addition, we will iteratively upgrade the emotion intervention optimization strategy in combination with the pain points of actual clinical diagnosis and treatment, promoting the deep adaptation of technology to clinical needs.

Conclusion

This paper proposes an upper-limb mirror rehabilitation training robot system that integrates emotional factors. The system uses machine vision to recognize the patient's emotional state and motion trajectory on the unaffected side while incorporating emotional factors into the rehabilitation process to optimize their rehabilitation experience. The patient's facial expressions and motion trajectory on the unaffected side are visually observed using this system. By analyzing emotional states through the system's emotion recognition algorithm as a monitoring indicator, rehabilitation therapists can identify and analyze the patient's psychological state and then design personalized physiological and psychological intervention plans. In the future, we will incorporate physiological electrical signals as assessment indicators into the motion intention recognition and emotion recognition modules to improve the system's accuracy in motion intention and emotion recognition. Clinical trials will also be conducted to refine the emotion-integrated rehabilitation optimization strategy and enhance rehabilitation outcomes.

Ethical approval

This study has passed the review of the Ethics Committee of Nanjing Qixia District Hospital (Report No. 20220609), and the Ethics Committee approval date is June 9, 2022.

Consent for publication

The consent for publication of this article has been obtained from the volunteers who participated in the experiment.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article. This work was supported in part by the National Key Research and Development Program of China (2022YFC2405600), in part by the National Natural Science Foundation of China (62075098, 61401217, W2433193), in part by the Leading-edge Technology and Basic Research Program of Jiangsu (BK20192004D), in part by the Key Research and Development Program of Jiangsu (BE2022160), and in part by the Natural Science Foundation of Jiangsu Province China (BK20230301). This work was also supported by Major Project of Basic Science (Natural Science) Research in Higher Education Institutions of Jiangsu Province (25KJA416001).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

The raw data analyzed in this study are publicly available from the AffectNet dataset,³⁵ accessible via its DOI: 10.1109/TAFRC.2017.2740923.

ORCID iDs

Shijie Liu  <https://orcid.org/0009-0006-2839-3338>
 Zhen Chen  <https://orcid.org/0000-0003-2588-5610>
 Lijun Mao  <https://orcid.org/0009-0005-9679-235X>
 Wenyan Li  <https://orcid.org/0009-0004-9265-7620>
 Shi Jie Wang  <https://orcid.org/0009-0008-2250-7829>
 Ziyang Zhou  <https://orcid.org/0009-0001-4769-7453>
 Yixue Liu  <https://orcid.org/0009-0005-7072-1318>
 Yumeng Yang  <https://orcid.org/0009-0000-8649-0650>
 Yilu Kong  <https://orcid.org/0009-0002-9170-0493>
 Keshu Cai  <https://orcid.org/0000-0002-9475-8343>
 Jianqing Li  <https://orcid.org/0000-0002-3524-8933>
 Liang Li  <https://orcid.org/0000-0002-0417-8995>
 Bin Liu  <https://orcid.org/0000-0003-3477-3178>

References

1. Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization (WSO): global stroke fact sheet 2022. *Int J Stroke* 2022; 17: 18–29.
2. Feigin VL, Owolabi MO, Abd-Allah F, et al. Pragmatic solutions to reduce the global burden of stroke: a world stroke organization–lancet neurology commission. *Lancet Neurol* 2023; 22: 1160–1206.
3. Cirstea CM. Integrating movement quality in stroke neuro-rehabilitation could enhance arm recovery. *Stroke* 2024; 55: 1098–1100.

4. Chang WH and Kim Y-H. Robot-assisted therapy in stroke rehabilitation. *J Stroke* 2013; 15: 174.
5. Lencioni T, Fornia L, Bowman T, et al. A randomized controlled trial on the effects induced by robot-assisted and usual-care rehabilitation on upper limb muscle synergies in post-stroke subjects. *Sci Rep* 2021; 11: 5323.
6. Gandhi DB, Sterba A, Khatter H, et al. Mirror therapy in stroke rehabilitation: current perspectives. *Ther Clin Risk Manag* 2020; 16: 75–85.
7. Covaciu F, Gherman B, Vaida C, et al. A combined mirror–EMG robot-assisted therapy system for lower limb rehabilitation. *Technologies* 2025; 13: 227.
8. Lum PS, Van der Loos M, Shor P, et al. A robotic system for upper-limb exercises to promote recovery of motor function following stroke. In: *ICORR'99: International Conference on Rehabilitation Robotics*, Stanford, CA, 1999.
9. Cui J-W, Li Z-G, Du H, et al. Recognition of upper limb action intention based on IMU. *Sensors* 2022; 22: 1954.
10. Wang C, Peng L and Hou Z-G. A control framework for adaptation of training task and robotic assistance for promoting motor learning with an upper limb rehabilitation robot. *IEEE Trans Syst Man Cybernet Syst* 2022; 52: 7737–7747.
11. Tappeiner L, Ottaviano E and Husty ML. A cable-driven robot for upper limb rehabilitation inspired by the mirror therapy. In: Zeghloul S, Romdhane L and Laribi MA (eds) *Computational kinematics*. Cham: Springer International Publishing, 2018, pp.174–181.
12. Gu P, Ding Y, Ruchi M, et al. Post-stroke dizziness, depression and anxiety. *Neurol Res* 2024; 46: 466–478.
13. Ghazali AS, Sidek SN and Fatai S. Development of emotional state model using electromagnetic signal information for rehabilitation robot. *Int J Comput Intell Syst* 2016; 9: 65–79.
14. Pandey AK and Gelin R. A mass-produced sociable humanoid robot: Pepper: the first machine of its kind. *IEEE Robot Autom Mag* 2018; 25: 40–48.
15. Lu H, Schomaker LR and Carloni R. IMU-based deep neural networks for locomotor intention prediction. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp.4134–4139. IEEE.
16. Zhang P, Zhang J and Elsabbagh A. Lower limb motion intention recognition based on sEMG fusion features. *IEEE Sensors J* 2022; 22: 7005–7014.
17. Chen Y, Zhang H, Wang C, et al. A hierarchical dynamic Bayesian learning network for EMG-based early prediction of voluntary movement intention. *Sci Rep* 2023; 13: 4730.
18. Dong F, Wu L, Feng Y, et al. Research on movement intentions of human's left and right legs based on electroencephalogram signals. *J Med Devices* 2022; 16: 041012.
19. Sarhan SM, Al-Faiz MZ and Takhakh AM. A review on EMG/EEG based control scheme of upper limb rehabilitation robots for stroke patients. *Heliyon* 2023; 9: e18308.
20. Gümüşlü E, Erol Barkana D and Köse H. Emotion recognition using EEG and physiological data for robot-assisted rehabilitation systems. In: *Companion Publication of the 2020 International Conference on Multimodal Interaction*, 2020, pp.379–387.
21. Ismail SNMS, Aziz NAA and Ibrahim SZ. A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG). *J King Saud Univ Comput Inf Sci* 2022; 34: 3539–3558.
22. Chanel G, Kronegg J, Grandjean D, et al. *Emotion assessment: arousal evaluation using EEG's and peripheral physiological signals*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp.530–537.
23. Kedilioglu O, Bocco TM, Landesberger M, et al. ARUCOE: enhanced ArUco marker. In: *2021 21st International Conference on Control, Automation and Systems (ICCAS)*, 2021, pp.878–881. IEEE.
24. Khazetdinov A, Zakiev A, Tsoy T, et al. Embedded ArUco: a novel approach for high precision UAV landing. In: *2021 International Siberian Conference on Control and Communications (SIBCON)*, 2021, pp.1–6. IEEE.
25. Poroykov A, Kalugin P, Shitov S, et al. Modeling ArUco markers images for accuracy analysis of their 3D pose estimation. In: *30th international conference on computer graphics and machine vision (GraphiCon 2020) Part*, 2020.
26. Park FC and Martin BJ. Robot sensor calibration: solving $AX=XB$ on the Euclidean group. *IEEE Trans Robot Autom* 1994; 10: 717–721.
27. Tsai RY and Lenz RK. A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration. *IEEE Trans Robot Autom* 1989; 5: 345–358.
28. Fuhrmann AL, Splechtna R and Přikryl J. Comprehensive calibration and registration procedures for augmented reality. In: *Immersive Projection Technology and Virtual Environments 2001: Proceedings of the Eurographics Workshop in Stuttgart*, Germany, May 16–18, 2001, pp.219–227. Springer.
29. Zhao C, Wei Y, Xiao J, et al. Inverse kinematics solution and control method of 6-degree-of-freedom manipulator based on deep reinforcement learning. *Sci Rep* 2024; 14: 12467.
30. Kucuk S and Bingul Z. *Robot kinematics: Forward and inverse kinematics*. London, UK: INTECH Open Access Publisher, 2006.
31. Xu R, Wang P, Li X, et al. YOLO-ARGhost: a lightweight face mask detection model. *J Supercomput* 2024; 80: 3162–3182.
32. Khanam R and Hussain M. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:241017725* 2024.
33. Mandal B, Okeukwu A and Theis Y. Masked face recognition using resnet-50. *arXiv preprint arXiv:210408997* 2021. DOI: 110.48550/arXiv2104.08997.
34. Yang S, Luo P, Loy C-C, et al. Wider face: a face detection benchmark. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp.5525–5533.
35. Mollahosseini A, Hasani B and Mahoor MH. Affectnet: a database for facial expression, valence, and arousal computing in the wild. *IEEE Trans Affect Comput* 2017; 10: 18–31.