

Research Article

Artificial Intelligence Assistive Technology in Hospital Professional Nursing Technology

Yanxue Cai,¹ Moorhe Clinto,² and Zhangbo Xiao ¹

¹Second Affiliated Hospital, Qiqihar Medical University, Qiqihar 161006, China

²Cuyahoga Community College, Cleveland, Ohio 43085, USA

Correspondence should be addressed to Zhangbo Xiao; xzbmd@qmu.edu.cn

Received 28 August 2021; Revised 9 October 2021; Accepted 15 October 2021; Published 27 November 2021

Academic Editor: Kalidoss Rajakani

Copyright © 2021 Yanxue Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Global aging is becoming more and more serious, and the nursing problems of the elderly will become very serious in the future. The article designs a control system with ATmega128 as the main controller based on the function of the multifunctional nursing robot. The article uses a convolutional neural network structure to estimate the position of 3D human joints. The article maps the joint coordinates of the colour map to the depth map based on the two camera parameters. At the same time, 15 joint heat maps are constructed with the joint depth map coordinates as the centre, and the joint heat map and the depth map are bound to the second-level neural network. The prediction of the position of the user's armpit is further completed by image processing technology. We compare this method with other attitude prediction methods to verify the advantages of this research method. The research background of this article is carried out in the context of global aging in the 21st century.

1. Introduction

In recent years, the problems of walking and movement of the elderly, the disabled, and others have received more attention from scientific researchers in various countries. Many countries have carried out scientific research in this field. Most research studies on human pose recognition focus on estimating 2D human joint coordinates from colour maps. The deep learning method based on large datasets has shown excellent results in detecting joint human points in colour images [1]. However, these algorithms cannot directly provide the position of the human body in the global coordinate system for the transfer and transportation care robot.

To ensure the accuracy of human body posture recognition of the transfer and transportation nursing robot, this paper uses the colour map human joint detection model as the first-level neural network to calculate the colour map human joint pixel coordinates. The article maps the joint coordinates of the colour map to the depth map based on the two camera parameters. At the same time, 15 joint heat maps are constructed with the joint depth map coordinates as the

centre, and the joint heat map and the depth map are bound to the second-level neural network. Finally, this paper obtains the global coordinates of 15 joints. Artificial intelligence is a technology developed to simulate, extend, and expand human intelligence.

2. Transfer and Transportation Nursing Robot

Figure 1 shows the transfer and transportation nursing robot "Baize" developed by the team. The human-like back-carrying robot uses sound source localization, visual recognition, and other means to realize the position localization and posture recognition of the person cared for. The functions of autonomous obstacle avoidance and autonomous navigation are realized through the perception of the surrounding environment [2]. Tactile sensors are installed on the robot arm, chest rest, and seat to sense the user's back hug status in real time. The real-time safety guarantee module in the core control system adjusts the robot's movements and can safely and comfortably complete the actions of lifting, carrying, transferring, and placing. These designs provide intelligent transport services for groups with lower limb inconvenience.



FIGURE 1: The transfer and transportation care robot.

These all illustrate the application of artificial intelligence assistive technology in hospital professional nursing technology.

To achieve the above functions, the team designed the transfer and transportation nursing robot (Figure 2). The transfer nursing robot first uses the microphone array and the sound processing module to recognize the user's voice and calculates the user's angle relative to the robot to turn to the user. The human body gesture recognition module and the path planning module move to the user in front of the user at a short distance. Finally, the robot drives to the destination and places the user according to the guidance of the path plan.

3. Level 1 Convolutional Neural Network

According to the needs of the transfer and transportation nursing robot, this article defines the human joints as 15 joint points (Figure 3). The human body 3D joint position prediction is divided into a two-level network [3]. In the first-level network, we use PAF (part affinity field) to predict the state of the human body in the colour map. After calculating each pixel's human joint likelihood value, we take the maximum coordinate as the joint coordinate.

3.1. 2D Joint Point Detection. The PAF method has the characteristics of high precision and high robustness. It can effectively detect the 2D posture of the human body from the RGB image. This method uses a 2D human joint detection model to generate a multichannel scoring heat map $S_{\text{est}} \in \mathbb{R}^{n \times m \times 15}$, $S_{\text{est}} = (S_0, S_1, \dots, S_i, \dots, S_{14})$. We order $p_i = \arg \max_{(u_s, v_s)} S_i(u_s, v_s)$, where u_s, v_s represents the coordinates of the heat map and p_i represents the coordinates of the maximum score value in the heat map S_i , that is, the predicted pixel coordinates of the joint i . Then, we get the predicted coordinates $p_{\text{est}} = (p_0, p_1, \dots, p_i, \dots, p_{14})$, $p_{\text{est}} \in \mathbb{R}^{2 \times 15}$, of 15 joints. To further improve the adaptability of the PAE method to the working environment of nursing robots, we use the weights provided by Open Pose as the initial weights. At the same time, we use the family environment dataset to perform migration learning on the network [4].

3.2. Loss Function. To predict the joint pixel positions, we define a 15-channel real heat map during model training. S_{gt} represents 15 Gaussian images with the real joint pixel position as the extreme centre. We calculate the minimum mean square error of S_{gt} and S_{est} as the loss function of model optimization:

$$L_{2D} = \frac{1}{15} \sum_{(u_c, v_c, i)} \|S_{gt}(u_c, v_c, i) - S_{\text{est}}(u_c, v_c, i)\|. \quad (1)$$

4. Joint Heat Map

This paper constructs a multichannel joint heat map with the pixel coordinates of the 15 human joint point depth maps as the centre. We first calculate the depth map mapping relationship of the colour map to obtain the joint depth coordinates. The article obtains the denoising depth map:

$$I_{\text{depth}}(u_{\text{depth}}, v_{\text{depth}}) = \text{med}\{D(u_{\text{depth}} - k, v_{\text{depth}} - l)\}, k, l \in W, \quad (2)$$

where $(u_{\text{depth}}, v_{\text{depth}})$ represent the pixel coordinates of the depth map; k, l represent the values in the template W ; and med stands for median operation. Considering that the pixel coordinates of the depth map can be mapped to the camera coordinate system, this paper maps the I_{depth} pixel to the colour map I_{color} to obtain the two-picture pixel coordinate registration relationship map M_{reg} . In this way, the coordinates of the joint depth map can be calculated in reverse [5]. The process is as follows.

Step 1. The article obtains I_{depth} , I_{color} , and M_{reg} at the same time. Add 2 channels based on the colour image I_{color} . These channels are used to store the corresponding depth map pixel coordinates. Define the coordinate of the pixel x_{depth} in I_{depth} as $(u_{\text{depth}}, v_{\text{depth}})$.

Step 2. Construct a 3-dimensional vector $p_{\text{depth}} = (u_{\text{depth}}, v_{\text{depth}}, z)$ for pixel x_{depth} , where z is the pixel value of I_{depth} at x_{depth} .

Step 3. Calculate the coordinate p_{depth}^{\sim} of x_{depth} in the depth camera coordinate system through the depth camera internal parameter matrix H_{depth} and p_{depth} .

$$p_{\text{depth}}^{\sim} = H_{\text{depth}}^{-1} p_{\text{depth}}. \quad (3)$$

Step 4. The article uses the space coordinate transformation matrix to transform p_{depth}^{\sim} into the coordinate p_{color}^{\sim} in the colour camera coordinate system.

$$p_{\text{color}}^{\sim} = p_{\text{depth}}^{\sim} + T, \quad (4)$$

where R, T represent the rotation matrix and translation matrix between the two cameras, respectively.

Step 5. Calculate the corresponding coordinate p_{color} of x_{depth} in the colour image coordinate system through the

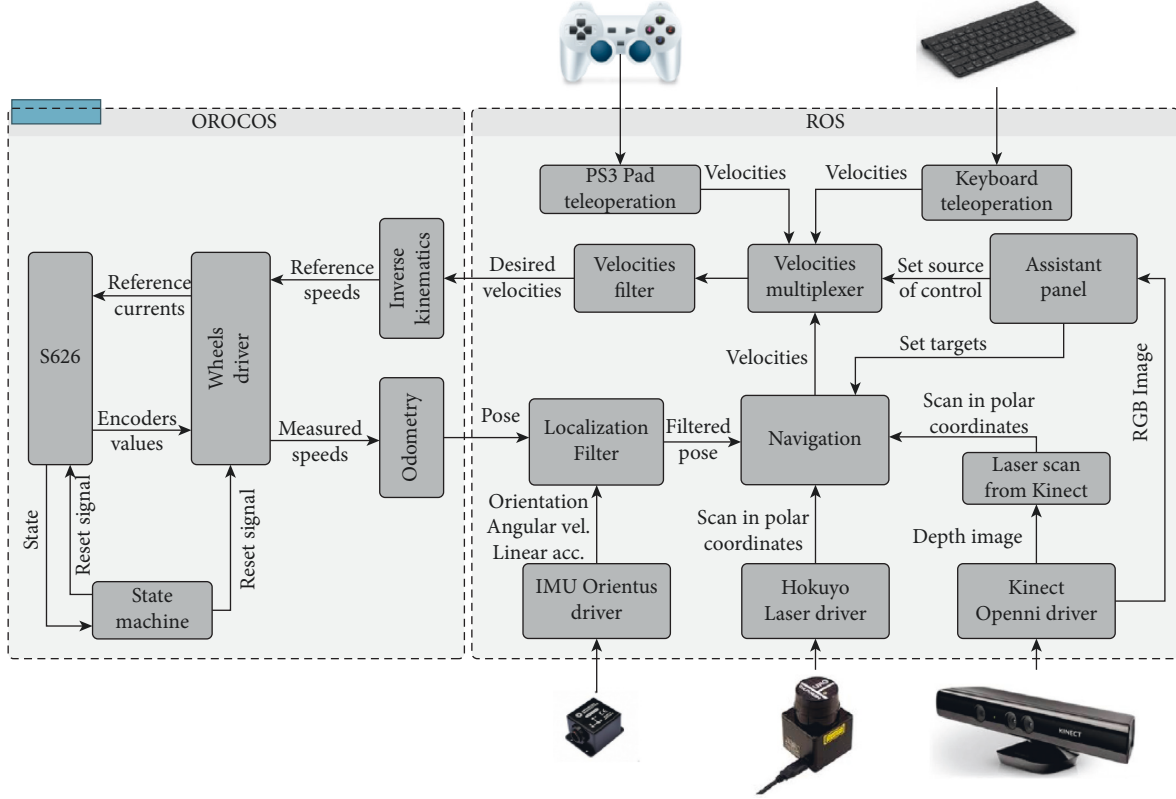


FIGURE 2: The control architecture of the transfer care robot.

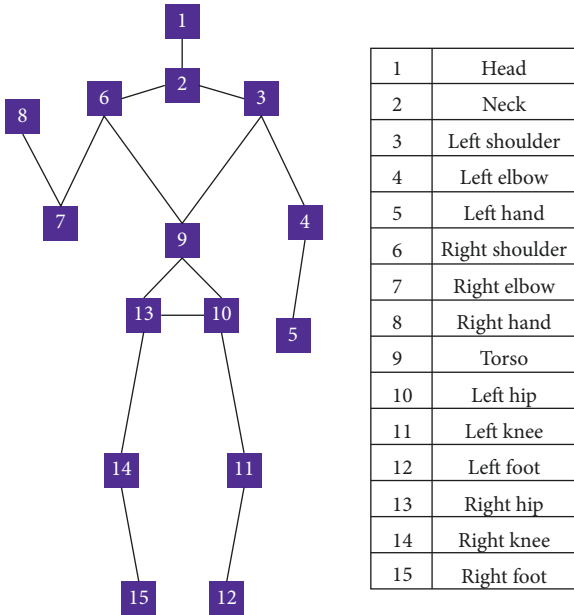


FIGURE 3: Distribution of human joints.

colour camera internal parameter matrix H_{color} and the colour camera coordinate system coordinate p_{color}^{\sim} .

$$p_{\text{color}} = H_{\text{color}} p_{\text{color}}^{\sim} \quad (5)$$

Repeat the above calculation for each pixel in I_{depth} in turn to obtain the coordinates of the mapped colour map. In this way, the pixel coordinates of the depth map corresponding to the pixel coordinates of each colour map can be obtained in reverse. Finally, we get the $R - G - B - U_{\text{depth}} - V_{\text{depth}}$ five-channel registration image M_{reg} .

Step 7. Find the joint depth map coordinate $p_{\text{est_depth}}$ corresponding to the joint colour map coordinate p_{est} according to M_{reg} . We define the multichannel joint heat map $H = (H_0, H_1, \dots, H_i, \dots, H_{14})$, with i as the joint index [6]. The H_i dimension of the heat map is the same as the I_{depth} dimension. $p_h = (u_h, v_h)$ is the pixel coordinate of the heat map. We use the 1-dimensional Gaussian function $h()$ to calculate H_i with the coordinate $p_{\text{est_depth}}$ of the joint depth map as the centre:

$$H_i(u_h, v_h, i) = g\left(\|p_h - p_{\text{est_depth}}(i)\|\right). \quad (6)$$

5. Level 2 Convolutional Neural Network

In the second stage, we use depth image and multichannel joint heat map binding as input. We further optimize the 2D joint detection results through convolutional neural networks to obtain 3D human joint poses. The algorithm flow is shown in Figure 4.

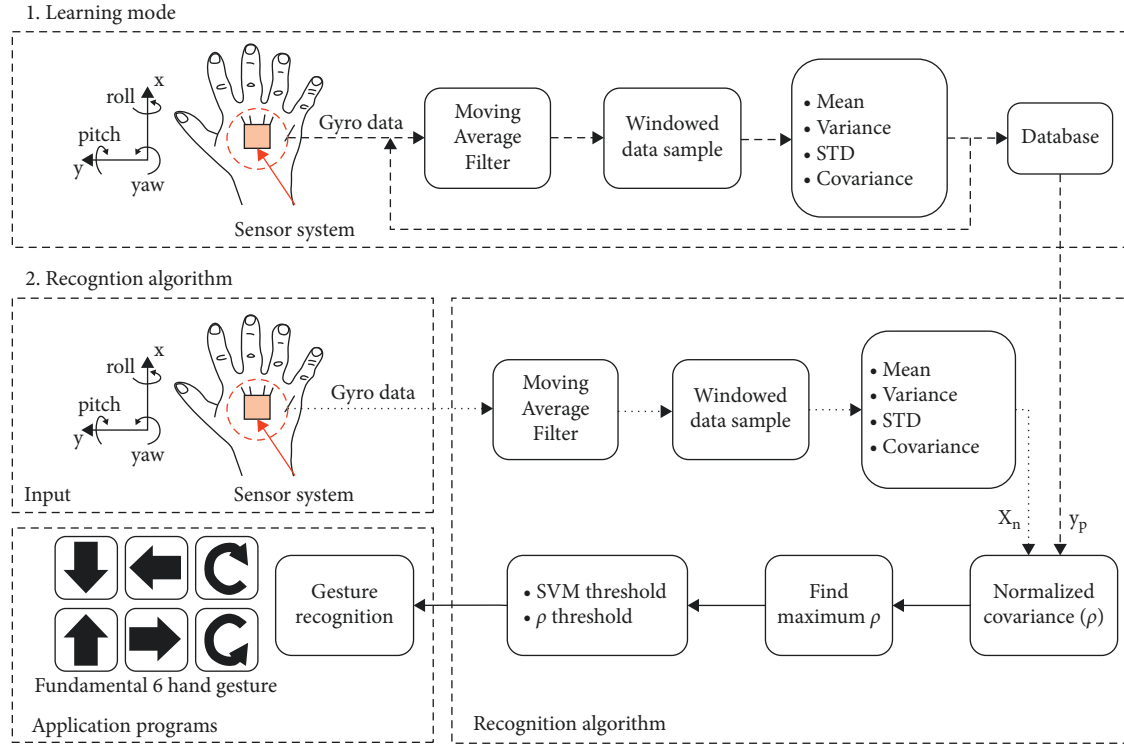


FIGURE 4: Human pose recognition algorithm.

5.1. Network Structure. The real-time human posture recognition algorithm based on a 3D convolutional network relies on a high-performance GPU (graphics processing unit), which is unsuitable for ordinary households' transfer and transportation care robot systems [7]. Therefore, to meet the accuracy requirements of joint prediction, especially near-range joint prediction, this paper uses a compromised 2D convolution method instead of 3D convolution. Based on the VGG16 network structure and the spatial pyramid pooling method (SPP), we design a convolutional neural network that is not limited by the size of the input image (Figure 5). We replace the 1000-unit fully connected layer final output by VGG16 with a 45-unit fully connected layer. The $u-v-z$ coordinates of 15 joints are estimated to get $P_{\text{est_depth}} = (P_{\text{est_depth}}^0, P_{\text{est_depth}}^1, \dots, P_{\text{est_depth}}^i, \dots, P_{\text{est_depth}}^{14})$, $P_{\text{est_depth}} \in \mathbb{R}^{3 \times 15}$.

Defining the global coordinates is the same as defining the depth camera coordinates. The depth map camera coordinates $P_{\text{est_depth}}^i$ are converted to the global coordinates $P_{\text{est_w}}^i$ by formula (7), $P_{\text{est_w}}^i = (x_w^i, y_w^i, z_w^i)$ is obtained, and then $P_{\text{est_w}} = (P_{\text{est_w}}^0, \dots, P_{\text{est_w}}^i, \dots, P_{\text{est_w}}^{14})$ is obtained.

$$\begin{cases} x_w = z_w \frac{(u - u_{\text{cam}})}{f_x}, \\ y_w = z_w \frac{(v - v_{\text{cam}})}{f_y}, \\ z_w = z. \end{cases} \quad (7)$$

Among them, $f_x, f_y, u_{\text{cam}}, v_{\text{cam}}$ represent the internal camera parameters and x_w, y_w, z_w represent the coordinate values of the global coordinate system.

In this paper, based on the ITOP (invariant-top view) dataset, the application scenario data training network of the transfer and handling nursing robot is added.

5.2. Loss Function. We use the minimum mean square error between the predicted joint global coordinates and the actual joint global coordinates as the loss function of the second-level neural network:

$$L_{3D} = \frac{1}{15} \sum_{i=0}^{i=14} \|P_{\text{est_w}}^i - P_{\text{gt_w}}^i\|, \quad (8)$$

where i is the joint index, $P_{\text{est_w}}^i$ is the predicted joint position, and $P_{\text{gt_w}}^i$ is the actual joint position.

6. Estimation of Underarm Points

We put the robot's two arms under the user's armpit which is the key to successfully picking up the user. This paper delineates ROI (region of interest) for the depth image based on the given depth image joint coordinates [8]. Then, we perform threshold segmentation in the ROI to obtain the underarm area and determine the target point where the robot arm is placed.

6.1. Pretreatment. We use the left (right) shoulder joint, left (right) elbow joint, and the midpoint of the left and right hip joints as vertices and delineate two triangles' ROIs. To ensure

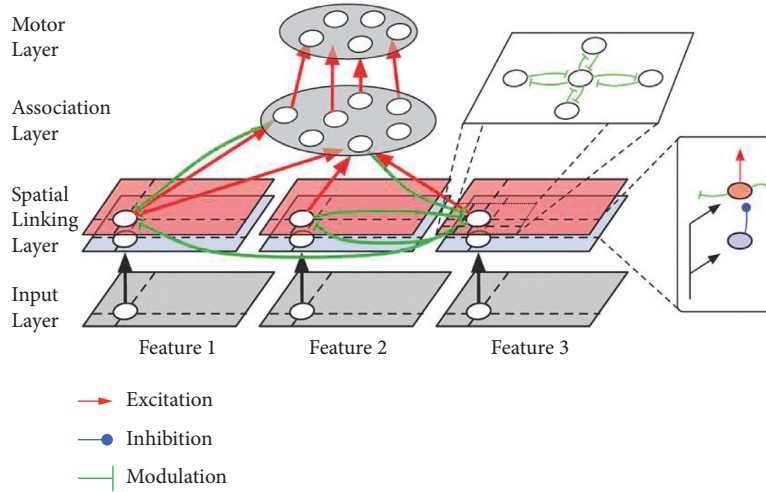


FIGURE 5: Level 2 neural network structure.

that the armpit area obtained by image segmentation does not contain the human body, we map the grey pixel value in the human body area to 1/3 of the original range [9]. We calculate the maximum depth j_{\max} and minimum depth j_{\min} among the four joint positions of the left (right) shoulder joint, left (right) elbow joint, left hip joint, and right hip joint in each ROI in turn. According to the pixel value mapping function shown in formula (9), the ROI is converted to grayscale.

$$\begin{cases} p_{\text{out}} = p_{\text{in}}, & p_{\text{in}} < j_{\min}, \\ p_{\text{out}} = \frac{1}{3}p_{\text{in}} + \frac{2}{3}j_{\min}, & j_{\min} \leq p_{\text{in}} < j_{\max}, \\ p_{\text{out}} = p_{\text{in}} - \frac{1}{3}j_{\max} + \frac{2}{3}j_{\min}, & j_{\max} \leq p_{\text{in}}. \end{cases} \quad (9)$$

6.2. Estimation of Axillary Points. We first perform 5×5 median filter denoising processing on the depth map. Considering that the background and foreground depth values vary with the scene, this paper uses the Otsu segmentation algorithm based on the maximum between-class variance to obtain the mask image M of the armpit area [10]. Further, we calculate the pixel coordinate $p_{\text{est_sp}}$ of the armpit centre point by the following formula:

$$p_{\text{est_sp}} = \frac{\sum_{(i,j)} M(i,j) \times (i,j)}{\sum_{(i,j)} M(i,j)}. \quad (10)$$

7. Experiments and Results

The PC configuration used in this experiment is Intel Core i7-6700HQ CPU, with 8G RAM and 4G NVIDIA GeForce GTX 950M GPU. The colour map and depth map sensors use the Microsoft Kinect 2 sensor. For the lower limb inconvenience group, we selected 1,100 human body images in the home environment as the test dataset.

7.1. Real-Time Evaluation. The experimental results of the calculation speed of commonly used gesture recognition methods are shown in Table 1. The comparison shows that the 3D convolution method used in [10] single-time 3D joint position estimation exceeds elapsed time 2s. This cannot meet the real-time requirements of the transfer and handling nursing robot. Therefore, the follow-up experiments of this article will not be tested. The Kinect SDK v2.0 method can complete about 30 joint predictions per second. The real-time performance of the algorithm at this time is optimal. The method in [9] is similar to the method in a single calculation, and the time is about 200 ms. This can prove the good real-time performance of the method in this paper.

7.2. Average Accuracy Evaluation. We use the 10 cm rule as the evaluation criterion for joint prediction. The article adopts the average accuracy of joint prediction (mAP) to measure the accuracy of the method used. Under the condition that the distance between the user and the camera is 550 mm–3500 mm, we collect 600 pieces of actual work scene data of the transfer and transportation nursing robot. The results of our evaluation of Kinect SDK v2.0, the method of Fujisawa et al. [9], and this paper's method are shown in Figure 6.

It can be seen from Figure 6 that when we predict each joint of the human body, the average estimation accuracy of the KinectSDKv2.0 method is the lowest. The direct mapping method used in [9] has good environmental adaptability and high accuracy. However, due to the colour map, the joint estimation itself has certain errors. This method easily predicts failure in relatively slender parts such as hands, elbows, and knees. In addition, the method is also deeply affected by the invalid value of the depth map and the noise, which reduces the accuracy. Compared with the above methods, this method has the highest accuracy, and the average accuracy of joints reaches 91.5%.

7.3. Accuracy Evaluation at Close Range. We use the transfer and transportation care robot developed by our team. We collect 350 working environment datasets and evaluate the

TABLE 1: Operation time for gesture recognition.

	Kinect SDK v2.0	[9]	[10]	Method of this article
Single calculation time (ms)	33	180	2400	210

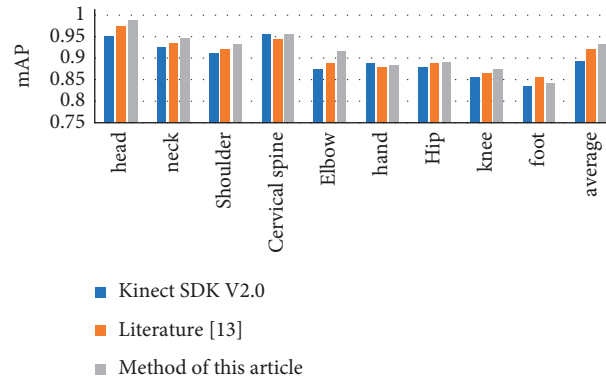


FIGURE 6: Comparison of average estimation accuracy.



FIGURE 7: Example of axillary point prediction.

close-range robustness of the method in this paper under the short-distance conditions with the operating range of 550 mm–800 mm. This article only evaluates the estimation effect of each method on the near-distance joint estimation based on the coordinate prediction of the head, neck, left and right hips, left and right elbows, and left and right shoulders.

It can be seen from Figure 7 that all three methods have reached the highest estimation accuracy on the head. Kinect SDK v2.0 and [9] methods have the lowest estimation accuracy at the hip joint, while the method in this paper has the lowest estimation accuracy at the elbow joint. Overall, the estimation accuracy of this method in any joint is the highest among the three methods. The average accuracy of the joint reaches 90.3%. When the distance is between 550 mm and 800 mm, the human body information collected by the depth camera is not comprehensive. Compared with the average accuracy evaluation experiment, the accuracy of Kinect SDK v2.0 dropped sharply in the close-range experiment. The algorithm has misjudgement of joint positions and even misidentified joints that do not exist in the figure, which is unacceptable for the transfer and transportation nursing robot.

On the other hand, the depth camera based on the ToF (time of flight) principle is more sensitive to the wrinkles and colours of clothes under close-range conditions. In this way, it is easier to collect invalid values, which increases the probability of the direct mapping method [9] to map invalid values. This leads to an increase in the recognition error rate. In comparison, the two-stage network method proposed in this paper has better short-range adaptability.

7.4. Evaluation of the Accuracy of the Axillary Point Prediction.

This article defines the 3 cm non-background rule as the basis for evaluating the accuracy of the axillary points. The predicted axillary point cannot be located on the human body, and the distance from the real axillary point in the X-Y plane is not more than 3 cm. It is regarded as an accurate prediction. Before calculating the user's armpit position, the transfer and transportation care robot moves to the front of the human body to face the front of the human body. We collect 150 frontal human body images at a distance of 550 mm–800 mm. Based on this, the experimental results of the evaluation of the method in this paper are shown in Figure 7.

After calculation, the method's accuracy for predicting the axillary points in this paper has reached 91.3%. This is because most of the operating objects are elderly people with mobility impairments under the conditions of transfer and transportation care applications. The human body is mostly in a passive state without too many complicated movements, and the body's posture is relatively simple. In addition, the method in this paper is based on human posture detection to obtain ROI. The prediction accuracy of the axillary points is directly related to the recognition ability of left (right) shoulder joint coordinates, left (right) elbow joint coordinates, and left (right) hip joint coordinates. Therefore, when the camera is too close to collect specific joint information, the armpit point estimation cannot be reliably completed when the joint position is incorrectly recognized.

8. Conclusion

This paper designs a 3D human pose estimation system for home care robots. The system uses Kinect 2 as the RGB-D data acquisition device to realize human joint position reasoning based on colour map 2D human body pose estimation. Through image processing technology, the prediction of the position of the user's axillary point is further completed. We compare this method with other posture prediction methods to verify the advantages of this research method and illustrate the correctness of the application of artificial intelligence assistive technology in hospital professional nursing technology [11, 12].

Data Availability

The human body images data used to support the findings of this study are restricted by the Ethical Care Committee of Qiqihar Medical University in order to protect patient privacy. Data are available from Zhangbo Xiao (e-mail address: xzbmd@qmu.edu.cn) for researchers who meet the criteria for access to confidential data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the 2021 Science and Technology Projects of Basic Scientific Research Business Expenses of Provincial Universities in Heilongjiang Province (Grant 2021-KYYWF-0031) "Research on Clinical Nursing of Elderly Disabled Patients Based on Artificial Intelligence Assisted Technology." The authors would also like to thank Chuan Zuo, Sitong Ge and Hezhang Sha for foundation assisting in this study.

References

- [1] T. R. Clancy, "Artificial intelligence and nursing: the future is now," *The Journal of Nursing Administration: The Journal of Nursing Administration*, vol. 50, no. 3, pp. 125–127, 2020.
- [2] G. H. Jeong, "Artificial intelligence, machine learning, and deep learning in women's health nursing," *Korean Journal of Women Health Nursing*, vol. 26, no. 1, pp. 5–9, 2020.
- [3] A. Barrera, C. Gee, A. Wood, O. Gibson, D. Bayley, and J. Geddes, "Introducing artificial intelligence in acute psychiatric inpatient care: qualitative study of its use to conduct nursing observations," *Evidence-Based Mental Health*, vol. 23, no. 1, pp. 34–38, 2020.
- [4] T. Risling and C. Low, "Advocating for safe, quality and just care: what nursing leaders need to know about artificial intelligence in healthcare delivery," *Canadian Journal of Nursing Leadership*, vol. 32, no. 2, pp. 31–45, 2019.
- [5] L. Pruinelli and M. Michalowski, "Toward an augmented nursing-artificial intelligence future," *CIN: Computers, Informatics, Nursing*, vol. 39, no. 6, pp. 296–297, 2021.
- [6] D. Watson, J. Womack, and S. Papadakos, "Rise of the robots," *Critical Care Nursing Quarterly*, vol. 43, no. 3, pp. 303–311, 2020.
- [7] A. G. Peirce, S. Elie, A. George, M. Gold, K. O'Hara, and W. Rose-Facey, "Knowledge development, technology and questions of nursing ethics," *Nursing Ethics*, vol. 27, no. 1, pp. 77–87, 2020.
- [8] R. Fuller and A. Hansen, "Disruption a," *Nursing Administration Quarterly*, vol. 43, no. 3, pp. 212–221, 2019.
- [9] S. Fujisawa, M. Takaiwa, Y. Hirata, S. Kotosaka, and D. Chugo, "Special issue on nursing robots and support systems for welfare sites," *Journal of Robotics and Mechatronics*, vol. 33, no. 4, pp. 711–712, 2021.
- [10] T. Tanioka, Y. Yasuhara, M. J. S. Dino, Y. Kai, R. C. Locsin, and S. O. Schoenhofer, "Disruptive engagements with technologies, robotics, and caring," *Nursing Administration Quarterly*, vol. 43, no. 4, pp. 313–321, 2019.
- [11] H. Bao, H. Bi, X. Zhang et al., "Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: a multicenter, clinical-based, observational study," *Gynecologic Oncology*, vol. 8, no. 1, pp. 125–128, 2020.
- [12] C. X. Li, W. M. Fei, C. B. Shen et al., "Diagnostic capacity of skin tumor artificial intelligence-assisted decision-making software in real-world clinical settings," *Chinese Medical Journal*, vol. 8, no. 20, pp. 72–75, 2020.