Dean&Francis
ISSN 2959-409X

Interaction Method of Rehabilitation Robot Based on VPE and VEP

Zhiming Huang

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, 200030 China
*Corresponding author: SBC-23-8130@sbc.usst.edu.cn

Abstract:
The development of rehabilitation robots has shown significant potential in improving patient recovery through intelligent interaction mechanisms. The passage introduces two telerehabilitation methods called Vision-based Pose Estimation and Vision-based Environmental Perception. The first method is to apply a non-invasive solution in order to treat upper limbs by wearing a hand exoskeleton with human robot interaction. With a set of master-slave setup, the RGB-D camera captures an image of what operators perform and delivers it to the wearable hand exoskeleton remotely by a bidirectional combination link, while an external sensor records the forces. It is convenient to estimate data and find a better solution. The second method is about an exoskeleton robot that is equipped with the ability of environmental perception. It is surrounded by visional sensors, a control system and critical technologies. Vision-based Pose Estimation (VPE) and Vision-based Environmental Perception (VEP) are testified to be promising in the prospective telerehabilitation robot. This paper aims to provide a new perspective on human-robot interaction in the field of neurorehabilitation.

Keywords: Telerehabilitation, Master-slave setup, RGB-D camera, Hand exoskeleton.

1. Introduction

In recent years, the number of patients being rehabilitation with a rehabilitation robot after suffering from various kinds of diseases like stroke and cerebral palsy is increasing gradually. As we all know, if people suffer a stroke, the functions of physiology will drop dramatically. The ability to react and balance is likely to fall. At the same time, the muscle spasticity will take place [1]. The experts in the field of robotics are fond of conducting research to solve problems with the help of a rehabilitation robot. The application of rehabilitation robots decreases the amount of work of doctors and nurses, obtaining recognition and a good reputation, which is proven in clinical practice [2]. In the process of rehabilitation, the interaction between patients and robots is much more important. It is an interaction system that involves a series of relative subjects including computer science, biomechanics, neuroscience and so on based on safety, functionality and effectiveness [3]. Hence, the objective of the study is to introduce two kinds of interaction meth-

ISSN 2959-409X

ods in order to assist patients, especially old people, to rehabilitate effectively rehabilitating.

Focusing on the rehabilitation of the upper limb, it is impossible for therapists in the hospital to take care of patients at home. Therefore, telerehabilitation plays an important role in the process of treatment, which reduces cost and time and combines patients with a robot in the interaction system. The telerehabilitation interaction system is equipped with capabilities of controlling hands and arms, evaluation and real-time subjective monitoring in the convenience of recording data of all terms, like exerting forces on the joints. After that, operators can analysis the results transmitted by the intelligent robot to come up with more excellent schemes [4]. Noticeably, it provides repetitive and beneficial tasks for users to do rehabilitation training in their daily lives [5].

What is introduced in the investigation are two methods called Vision-based Pose Estimation (VPE) and Vision-based Environmental Perception (VEP) that are passive rehabilitation methods. In the first part, there is a set of devices to complete the overall rehabilitation work in the form of a master-slave setup. The master-slave setup is composed of a wearable hand exoskeleton for duplicating actions and an RGB-D (Red, green, blue and depth) camera associated with assessing the 3D position of the operator's hand joints and allowing them to control remotely. What's more, how to connect the master and slave unit and how patients practice correctly are displayed. In the second part, the VEP method is based on the relationship between users, environment and exoskeleton. It is complicated to describe the interaction between the human-robot and the environment, because there are enormous algorithms and electrical signal interactions behind it. However, some compositions and restricted factors are illustrated in this passage. The significance of incorporating vision lies in prediction about the upcoming environment and estimation about gait planning.

2. Vision-based Pose Estimation

2.1 Composition of VPE

The VPE system includes a type of master-slave setup. The master unit is an RGB-D camera utilized to record, deliver and cope with hand orders information from operators. The slave unit, which is a hand exoskeleton, will receive real-time order information and drive patients to perform the same action as operators. Meanwhile, the sensorized object will take notes about the magnitude of forces held by patients and send feedback to operators by the master unit[6]. The RGB-D camera is Figure 1, the wearable hand exoskeleton is Figure 2.

Fig. 1 RGB-D camera [7]

Fig. 2 wearable hand exoskeleton [8]

2.1.1 The Master Unit

The master unit will apply random forest (RF) to estimate the hand movement of operators.RF is able to calculate all probabilities of all joint movements. The RGB-D camera is hung 50cm above the table to avoid self-occlusions of the operator's fingers. The RF classifier achieved by a laptop reads depth input stream from the camera at the rate of 30fps, which is the highest working frequency of the RGB-D camera. Only does the RF classifier cope with a hand's pixel, but the pixel of a plane like a desk can not be segmented by it. The pixel of the desk is deleted by the RANSAC algorithm. Once the pixel information of the hand is segmented, the RF classifier deals with it easily. This way is beneficial for daily training. Five of all joints will be retained to calculate the sequence main sequence orders, these parts are metacarpo-phalangeal (MCP), proximal and distal-interphalangeal (PIP and DIP) joints [6]. The model and images used for training the RF classifier are Figure 3.

Fig. 3 Images of RGB-D camera [6]

2.1.2 The Slave Unit

The slave unit is used to capture sensorized objects by hand exoskeleton and send information on forces to the master unit accurately. It is a type of mechatronic device that consists of a wearable hand exoskeleton, a remote actuation block driven by a cable-sheath system and a power unit. There are four degrees of motion (DOF): MCP, P-DIP, MC-IP and carpo-metacarpal (CMC). For the index finger, it is actuated by MCP, PIP and DIP. For the thumb, it is actuated under flexion/extension(f/e) of MCP, DIP and CMC. The overall process of exoskeleton motion is driven by a direct current motor through a bidirectional cable-sheath transmission[6].

The working theory of an external sensorized grasping object is related to an electric circuit. It is a rectangular resin block, the widest surface is covered by two pressure-sensitive pads based on photoelectric sensing technology for measuring the forces of human-robot interaction. In addition, by squeezing two silicone block-shaped hollow structures, patients can grasp the sensorized object. There is an LED emitter and a light receiver on the circuit board. When a silicone block-shaped hollow structure is squeezed, deformation will obstruct the light collected by the receiver and illustrate the ratio of voltage drop in each receiver of the corresponding LED emitter [6]. The previous report about the sensor properties figures out that the force of human-robot interaction is 0.16 N [9]. The pinch grasp and lateral grasp practice are Figure 4.

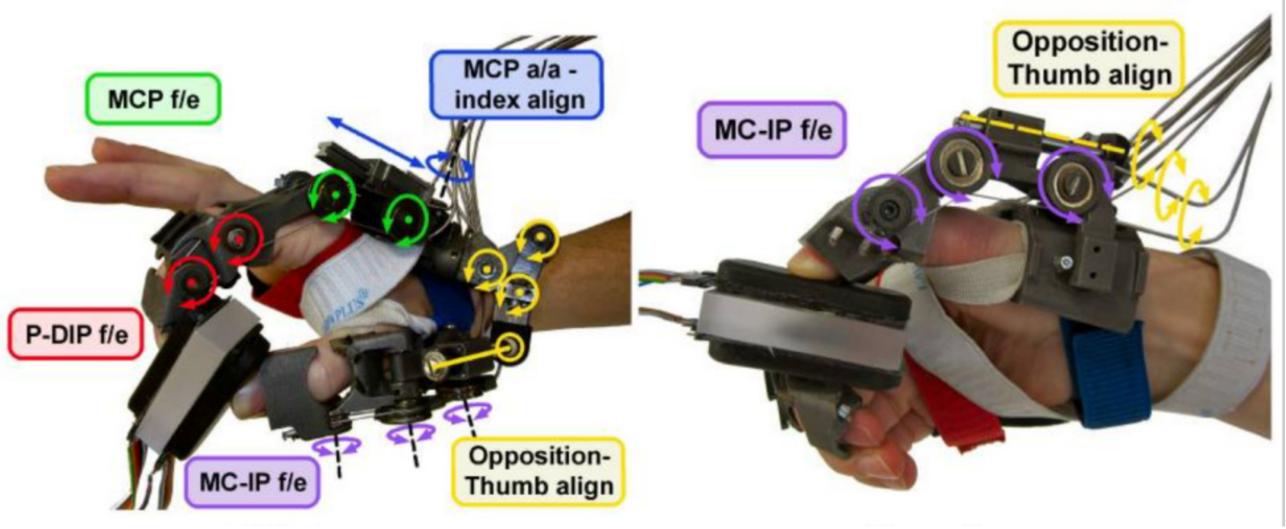

Fig. 4 Pinch grasp and lateral grasp practice [10]

In this paper, researchers Farulla and so on divided the practice process into two gestures: pinch grasp and lateral grasp. The VPE algorithm works by evaluating which joints the hand is interested and calculating the finishing percentage of practice. For pinch grasp, this percentage is linked to the standard distance between the index fingertip and thumb fingertip. For lateral grasp, this percentage is linked to the distance between the thumb fingertip and the plane normal along the MCP, PIP, DIP joints and index fingertip. It is because the velocity of calculation is fast [6]. The images of calculation about the finishing percentage of practice are Figure 5.

ISSN 2959-409X

Fig. 5 Images of calculation about finishing percentage of practice [6]

2.2 Communication between Master Unit and Slave Unit

What connects the master unit with the slave unit is a kind of bidirectional communication link working at the rate of 30 fps, which ensures consistency with the master unit. The master unit is used to encode tasks of expected motion data that is sent to the slave unit within a single byte. On the contrary, the slave unit can transfer the kinematic and dynamic states and forces of the sensorized object to the master unit. All of the operations are real-time (RT) actions [6].

3. Vision-based Environmental Perception

3.1 Selection and Installation of Visual Sensors

There are a variety of types of visual sensors. Selecting a visual sensor, researchers have to take size, weight, power consumption, performance, detection range, accuracy, visual angle, stability and compatibility into consideration. Every brand has its own unique visual identity on the market. At the same time, it is vital for researchers to figure out which part of the body the visual sensor is supposed to be installed exactly [7]. The requests for visual sensors are as follows. First, it is as small and light as possible within low power consumption in order to decrease the load of operation. Secondly, the application scenarios of it are able to adapt to different lighting conditions. Thirdly, in the aspect of software, the Application Programming Interface (API) offered by visual sensors must be consistent with mainstream edge computing platforms and programming languages. Fourthly, cost is a factor that can not be ignored. After all, nobody wants to buy an expensive product to treat the lower limb [11].

3.1.1 Classification of Visual Sensors

The visual sensors are divided into two kinds: passive sensors and active sensors according to whether they can emit an energy source into the environment [12].

The typical passive sensor is an RGB camera.RGB camera has been prevalent in people's daily life because of its advantages of low cost and small size. Nevertheless, RGB fails to absorb sufficient depth information from the environment. The working theory of an RGB camera is triangular measurement, meaning that it takes a photo of an object from three different views and calculates depth through disparity in different photos. The process is influenced by lighting conditions. That is why an RGB camera is constrained to capture depth information [13].

The principle of an active sensor indicates that the sensor emits a signal to the environment and reacts to the reflected signal to measure distance. The common active sensor in daily life is Time-of-Flight (TOF). Even if TOF has the features of wide range measurement, it can not go through without consuming excessive power all the time. A TOF camera obtains the right distance by measuring the time of light being emitted and reflected by Light Detection and Ranging (LiDAR). LiDAR exploits a rotating photoelectric diode to acquire a whole view of the environment [14].

With the development of the Internet and technology and deep investigation, scientists combine an active sensor with an RGB camera to create an RGB-D camera.RGB-D camera is equipped with the capacity to attain useful depth information of streets and roads under Micro-Electric-Mechanical System (MEMS) [15]. Some common visual sensors are as follows. The depth camera is Figure 6. The LiDAR camera is Figure 7. The stereo camera is Figure 8.

Fig. 6 Depth camera [16]

Fig. 7 LiDAR camera [17]

Fig. 8 Stereo camera [18]

3.1.2 Location of Visual Sensors

The accuracy and range of measurement depend on the location of the visual sensor. Sometimes, it is set on the patient's head, chest and lower limbs [19]. The advantages and shortcomings of different positions are as following Table 1 [7]. It can be seen from Table 1, the best positions of visual sensor installation are the chest and waist. These two points allow the horizon to be steady.

Table 1. Advantages and disadvantages of different positions

Installation Location	Advantages	Disadvantages	Suitable Devices
Head	Synchronizes with user's view	Heavy weight may lead to discomfort and shaky images	Blind guidance equipment,upper-limb exoskeletons
Chest	The images are stable, and the view is synchronized with movement	Camera posture is easily affected by upper-body movement	Upper-limb exoskeletons, lower-limb exoskeletons
Waist	The images are the most stable, and the view is synchronized with movement	Low field of view, limited visual range	Lower-limb exoskeletons, lower-limb prosthetics
Lower limb	High accuracy in detecting specific terrains at close range	Restrictions on users' lower-body dress, shaky images	Lower-limb prosthetics
Feet	High accuracy in detecting specific terrains at close range	Limited field of view, shaky images	Lower-limb prosthetics, smart shoes

3.2 Control System

To achieve exact perception in plentiful complicated circumstances, there is a powerful control system that tackles information from visual sensors and converts it into orders of relative motions. The control system is composed of three layers: high-level controller,mid-level controller and low-level controller [20].

A high-level controller inside the control system is assigned to pick up information from all of the sensors and encodes it so that the controller anticipates possible movements of the robot. Deep learning is becoming a pretty trend gradually, and it is integrated into high-level controllers to operate convolutional neural networks (CNN) effectively [7].

After the tasks of the high-level controller, the mid-level controller will generate a series of corresponding motion models. It is likely to be helpful to simulate all possible motion trails [7]. What's more, the real-time nature is what the mid-level controller must equip to transmit relative signals.

Eventually, the lower-level controller is responsible for adjusting various parameters, such as velocity, time and acceleration. Usually, the internal algorithm is Proportional-Integral-Derivative (PID) [7].

The whole operation procedure of the three controllers is as Fig. 9 [7].

ISSN 2959-409X

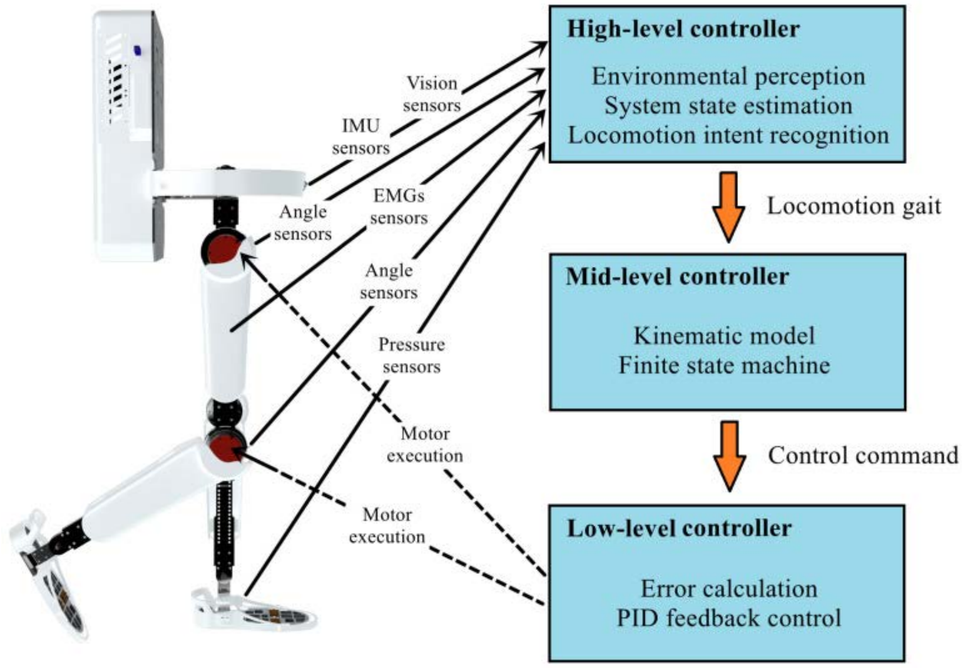

Fig. 9 Operation procedure of three controllers [7]

3.3 Analysis of Critical Technology

The critical technology lying in environmental perception of lower-limb exoskeleton contains classification of environments, related datasets, and gait planning facing a special environment. Analysis of these sides is beneficial to learn theories behind the lower-limb exoskeleton. First of all, the classification of environments has a much greater impact on subsequent algorithms to a certain degree. Some of the research relies on a multi-modal-fusion-based method. For example, Zhu studied the interaction mechanism between a flexible exoskeleton and the natural environment [21]. Besides, deep learning focuses on related datasets, which means that deep learning is driven by data [7]. Thirdly, gait planning facing a special environment is closer to planning about kinematics of corresponding joints [7].

4. Discussion

4.1 Similarities and Differences

Regardless of VPE or VEP, there are some similarities and differences between them.

Both VPE and VEP have the potential to promote progress in the field of rehabilitation robotics. The most obvious characteristic is vision-based.VPE and VEP both make full use of RGB-D cameras and sensors to process complicated information and encode with much more complex algorithms. In terms of cost, the current materials and devices for the two methods are expensive for most poor patients.

However, differences between the two methods are more significant.VPE focuses on the upper-limb while VEP underlines the lower-limb. Moreover, VPE counts on a force sensor, and VEP is determined by a visual sensor. VPE evaluates postures according to probability from random forest, VEP estimates motion trails depending on analysis of related datasets and gait planning facing special environments. In addition, VPE uses a master-slave setup to control the wearable hand exoskeleton remotely, three controllers of VEP control system are independent. In application positions, it is appropriate to apply VPE to rehabilitate at home. VEP is operated on different kinds of streets and roads.

4.2 Future Outlook

Even though current research about vision interaction is rare, scientists around the world devote themselves to exploring something more valuable. Many elements being developed like deep learning, image disposal and algorithms accelerate the development of vision interaction. It will become a crucial spot in all fields one day, especially the treatment of stroke. It is a fact that vision interaction has a positive impact on human robot interaction and is more competitive in the robot market in the future.

5. Conclusion

In conclusion, rehabilitation robot has made certain progress in the field of vision interaction, it still needs more contributions and algorithm constructions to support its development. This passage attaches importance to two scientific rehabilitation methods: VPE and VEP disclose the status and roles of two systems in human-robot interaction. The passage discusses two methods from two angles: composition and applications. These two methods are both based on vision interaction, reflecting that vision interaction in the field of rehabilitation robotics is likely to become an effective and practical way. As for restricted elements, both of them lack relative gait planning datasets, optimization of human roles in the loop and sufficient funding. These conditions constrain the development of vision interaction. In the future, diversified industrials acquire much profit from such cutting-edge technology. It is changing a large number of industries potentially, especially wearable exoskeletons. Hence, there is a lot of work being conducted in terms of algorithms and mechanical materials optimization to prove its feasibility. Anyway, VPE and VEP can be regarded as a practical tool for people who are tortured by stroke and cerebral palsy to recover earlier.

References

[1] Andrade R L, Figueiredo J, Fonseca P, Vilas-Boas J P, Silva M T, Santos C P. Human-robot joint misalignment, physical interaction, and gait kinematic assessment in ankle-foot orthoses. Sensors, 2023, 24(1): 246.

[2] Wendong W, Hanhao L, Menghan X, Yang C, Xiaoqing Y, Xing M, Bing Z. Design and verification of a human-robot interaction system for upper limb exoskeleton rehabilitation. Medical Engineering & Physics, 2020, 79: 19–25.

[3] Beckerle P, Salvietti G, Unal R, Prattichizzo D, Rossi S, Castellini C, Hirche S, Endo S, Amor H B, Ciocarlie M, Mastrogiovanni F, Argall B D, Bianchi M. A human-robot interaction perspective on assistive and rehabilitation robotics. Frontiers in Neurorobotics, 2017, 11: 24.

[4] Placidi G, Di Matteo A, Lozzi D, Polsinelli M, Theodoridou E. Patient-therapist cooperative hand telerehabilitation through a novel framework involving the virtual glove system. Sensors, 2023, 23(7): 3463.

[5] Dalla Gasperina S, Roveda L, Pedrocchi A, Braghin F, Gandolla M. Review on patient-cooperative control strategies for upper-limb rehabilitation exoskeletons. Frontiers in Robotics and AI, 2021, 8: 745018.

[6] Airò Farulla G, Pianu D, Cempini M, Cortese M, Russo L O, Indaco M, Nerino R, Chimienti A, Oddo C M, Vitiello N. Vision-based pose estimation for robot-mediated hand telerehabilitation. Sensors, 2016, 16(2): 208.

[7] Wang C, Pei Z, Fan Y, Qiu S, Tang Z. Review of vision-based environmental perception for lower-limb exoskeleton robots. Biomimetics, 2024, 9(4): 254.

[8] Sandison M, Phan K, Casas R, Nguyen L, Lum M, Pergami-Peries M, Lum P S. HandMATE: Wearable robotic hand exoskeleton and integrated Android app for at-home stroke rehabilitation. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2020, 4867–4872.

[9] Donati M, Vitiello N, De Rossi S M M, Lenzi T, Crea S, Persichetti A, Giovacchini F, Koopman B, Podobnik J, Munih M, et al. A flexible sensor technology for the distributed measurement of interaction pressure. Sensors, 2013, 13(1): 1021–1045.

[10] Cempini M, Cortese M, Vitiello N. A powered finger–thumb wearable hand exoskeleton with self-aligning joint axes. IEEE/ASME Transactions on Mechatronics, 2015, 20(3): 705–716.

[11] Krausz N E, Hargrove L J. A survey of teleceptive sensing for wearable assistive robotic devices. Sensors, 2019, 19(23): 5238.

[12] Nelson M, MacIver M A. Sensory acquisition in active sensing systems. Journal of Comparative Physiology A, 2006, 192(6): 573–586.

[13] Zhu X, Li Y, Lu H, Zhang H. Research on vision-based traversable region recognition for mobile robots. Applied Research of Computers, 2012, 29(7): 2009–2013.

[14] Hall D S. High definition lidar system. U.S. Patent EP2041515A4, 2009.

[15] Intel RealSense. LiDAR Camera—Intel RealSense Depth and Tracking Cameras. 2013. https://www.intelrealsense.com/lidar-camera-l515/

[16] Intel RealSense. Depth Camera D435i—Intel RealSense Depth and Tracking Cameras. 2013. https://www.intelrealsense.com/depth-camera-d435i/

[17] Stereolabs. ZED Mini Stereo Camera. 2002. https://store.stereolabs.com/products/zed-mini

[18] Massalin Y, Abdrakhmanova M, Varol H A. User-independent intent recognition for lower limb prostheses using depth sensing. IEEE Transactions on Biomedical Engineering, 2018, 65(8): 1759–1770.

[19] Zhu H. Research on terrain recognition of flexible exoskeleton based on computer vision. Wuhan University of Technology, 2020.

[20] Krausz N E, Hargrove L J. A survey of teleceptive sensing for wearable assistive robotic devices. Sensors, 2019, 19(23): 5238.

[21] Zhu H. Research on terrain recognition of flexible exoskeleton based on computer vision. Wuhan: Wuhan University of Technology, 2020.