Youth physical training robot based on visual recognition and training method

Hanbin Dong*, Zhihe Cui, Liao Tian
Northeast Normal University, Changchun, Jilin 130500, China

*Corresponding author: 813492370@qq.com

Abstract. With the continuous improvement of people's living standards, the continuous development of training methods and concepts, resulting in the continuous improvement of sports level, which puts higher demands on teenagers' physical function. How to use high-efficiency and low-risk training methods to train has become a headache for coaches. In this paper, an accurate target detection and recognition method based on stereo matching algorithm is proposed. Based on stereo matching algorithm, after data training, the target can be accurately identified, and the position information of the target can be given. Then, the target can be accurately positioned by binocular vision and its three-dimensional coordinates can be obtained. Finally, it can be transmitted to the robot and control the robot to move and grab the target object. Experimental results show that this method can identify and locate the target accurately and quickly, and successfully guide the robot to grasp the target object. This physical training robot can provide theoretical basis and technical guidance for improving the physical training level and sports performance of young athletes.

Keywords: Visual recognition; Physical training for teenagers; Training robots; training method

1. Introduction
Physical fitness is the most basic exercise ability of human body. The majority of teenagers in the new period are not ideal in physical performance. Physical training is a process of improving body shape, improving the functions of various organs, fully developing sports quality, and promoting the improvement of physical exercise ability and sports performance by means of action exercises with reasonable load. Physical training should start with teenagers, build perfect physique from an early age, train strength quality, flexibility quality, speed quality, sensitivity quality and endurance quality, and train with different physical training equipment to achieve the effect of physical training for teenagers [1].

Robot-related technologies, such as motion control technology, computer vision technology, deep learning technology, and speech pickup semantic recognition technology, have made great progress in recent years, which has promoted the spurt development of the service robot industry. More and more service robots appear in life, such as sweeping robots, aerial drones, escort robots, home housekeeper robots, restaurant service robots, etc., making life more and more convenient [2-3]. Because there are
many kinds of objects in real environment with different shapes and sizes, the graspability marks of objects can only be obtained by manual marking according to the relationship between objects and manipulators, which makes it very difficult to collect marking data. Therefore, how to make the robot object recognition algorithm have certain generalization ability under the condition of insufficient training data, and obtain the graspability and graspable object types at the same time, is the research difficulty in this part.

The physical training robot based on visual recognition technology developed in this subject covers computer vision technology, mechanical design, motion control technology and so on, which has great research significance.

2. The purpose and significance of adolescent physical training

The improvement of physical function is obtained on the basis of physical practice, and in the process of physical practice, the body bears the energy demand of training through energy consumption, so there is a contradiction between the trainer's physical consumption and physical recovery. Generally speaking, the consumption of physical energy in sports training needs to be restored after sports training. There are obvious individual differences in this recovery process as far as the trained groups are concerned.

Therefore, it is not scientific to advocate and encourage hard training. Adolescence is a period when athletes lay the foundation, and the quality of its foundation is directly related to the future development of athletes. Teenagers repeat a kind of exercise for a long time in the process of exercise, which can easily lead to mental fatigue, thus causing nerve fatigue, and functional combination training on cushion is suitable for different groups of people. It can be perfectly combined with other training equipment to achieve the purpose of training, which can effectively stimulate people's interest in participating in sports, reduce the boring feeling of physical exercise, prevent the generation of nerve fatigue, and pad functional combination training with dumbbell, elastic belt, solid ball and other equipment combination training to better achieve the training effect.

3. Vision-based physical training robot

Robot needs to judge whether the objects in its field of view are suitable for grasping. There are many objects in human living environment that are not suitable for the robot manipulator to grasp. If the robot forcibly grasps these objects, it will lead to the failure of the grasping task and even cause damage to the manipulator. In this paper, the gripability of objects mainly focuses on the gripability of objects under visual observation, which is mainly determined by the relationship between the structure of robot manipulator and the size and shape of objects, without considering the influence of objects' weight and friction properties.

3.1. Imaging model and image correction of RGB-D camera

In the real environment, the robot's field of view will contain a lot of invalid background information, so it is necessary to segment the area where the object is located in the image first. Commonly used image segmentation methods mainly include: background removal method [4-5], image edge method and energy minimization method, etc. [6-7]. Among them, the background removal method needs to establish the background model of the environment, which is not suitable for still object segmentation, while the algorithm complexity of image edge method and energy minimization method is relatively high, which is not suitable for real-time image processing. Therefore, this paper designs a fast object segmentation method based on RGB-D information. The method uses 3D vision technology based on time-of-flight and optical coding technology to acquire RGB images and depth images of objects at the same time, and through the steps of object support plane fitting, "rough" segmentation of objects on depth images, "fine" segmentation of objects on color images and depth images, etc., the object region can be effectively segmented.

The imaging process of RGB camera can be described by three coordinate systems, namely, world coordinate system \((o_w, x_w, y_w, z_w)\), camera coordinate system \((o_c, x_c, y_c, z_c)\) and image coordinate
Among them, the image coordinate system is a two-dimensional coordinate system, the origin is located in the upper left corner of the camera imaging chip, and the $x$ axis and $y$ axis coincide with the length and width direction of the chip respectively; The origin of the camera coordinate system is located at the optical center of the lens, the $z$ axis is outward along the optical axis, the $x$ axis and $y$ axis are parallel to the $u$ axis and $v$ axis of the image coordinate system respectively, and the external world coordinate system is set as the base coordinate system of the robot.

Without considering lens distortion [8-9], the vision camera conforms to the pinhole imaging model, as shown in Figure 1.

![Aperture imaging model](image)

Figure 1 Aperture imaging model

It can be seen from fig. 1 that the coordinate point $(x_w, y_w, z_w)$ of the target in the world coordinate system, the point $(x_c, y_c, z_c)$ of the target in the camera coordinate system and the coordinate point $(x_v, y_v)$ in the image plane coordinate system satisfy equations (1) and (2). Equation (1) maps the point of the target in the world coordinate system to the camera coordinate system, and equation (2) maps the camera coordinate system to the image plane coordinate system.

$$
\begin{bmatrix}
x_c \\
y_c \\
z_c \\
1
\end{bmatrix}
H =
\begin{bmatrix}
x_w \\
y_w \\
z_w \\
1
\end{bmatrix}
(1)
$$

$$
\begin{bmatrix}
x_v \\
y_v
\end{bmatrix}
= \begin{bmatrix}
x_c \\
y_c \\
z_c \\
z_v
\end{bmatrix}
(2)
$$

Where $H$ is the homogeneous coordinate transformation matrix between the external world coordinate system and the camera coordinate system.

The RGB-D camera also includes a depth camera, which can directly obtain the distance information between each point in the field of view of the depth camera and the camera by measuring the phase delay of the transmitted and received infrared signals, and store it in the pixel of the image, which is very suitable for measuring the distance information in unstructured environment scenes.

3.2. Motion control module

The motion control module is responsible for executing the motion command of the upper computer, carrying the high-speed motion of the recognition module, and can be divided into two parts, namely hardware design and software design (Figure 2).
The hardware design includes the selection of single chip microcomputer, the selection of appropriate motor and motor driver according to the requirements of load and movement speed, the movement chassis of the robot, and related hardware circuits. The software design includes controlling the motor speed and compiling some specific motion control algorithms based on tennis court application.

3.3. Stereo matching
Stereo matching algorithms can be divided into two types: sparse point matching and dense point matching. Sparse point matching mainly aims at the strong feature points of the image, such as the contour, edge and corner points of the image, which express the essential attributes of the image and are obviously different from the surrounding pixels, so it is easier to determine and complete the matching. After strong feature point matching, sparse disparity map will be obtained, and dense disparity map needs to be obtained by interpolation. Dense point matching is to match all pixels in an image. In the process of dense point matching, because most areas of the image have no obvious features [10], and occlusion occurs, it is easy to happen mismatching, and the calculation of dense points is large, which requires higher hardware. Its advantage is that it can use all the information of the image, and then restore all the details of the object to the maximum extent.

BM (Block Matching) is a stereo matching algorithm which uses sparse point matching and detects similarity based on absolute error accumulation. BM algorithm only finds the strong matching points between two images for matching, and its operation speed is faster. The process is to first convert the image into a gray image, select a point on the reference image (usually the left image), intercept the square area of $n \times n$ with the point of interest as the center, and take a point on the same pixel line (through the previous stereo correction, the matching point must be in the same pixel line of Y coordinate) on the matching image (usually the right image), and also intercept this point as the center $n \times n$ square area, the gray similarity of the two square areas is detected. The points on the matching map are taken from left to right from the corresponding pixel rows according to the set disparity range, and the gray similarity of two squares is calculated one by one. The center points of the two square areas with the highest similarity are regarded as matching points, and the disparity is obtained.

The absolute error accumulation is a function for checking the gray similarity of two square areas of $n \times n$, and its mathematical expression is shown in formula (3).

$$C(x, y, d) = \sum_{x,y} |I_1(x, y) - I_2(x + d, y)|$$

$x, y$ refers to the coordinates of the feature points to be matched in the pixel coordinate system of the imaging plane, $S$ refers to an $n \times n$ square area centered on the feature points to be matched, $d$ refers to parallax, $I_1$ refers to the reference image, and $I_2$ refers to the matching image. $C(x, y, d)$ refers to the accumulated absolute value sum of gray scale differences between the corresponding points of the square area centered on the coordinate $(x, y)$ on the reference image $I_1$ and the square area centered on the coordinate $(x + d, y)$ on the matching image $I_2$ when the parallax is $d$. When the disparity $d$ reaches the maximum value from the minimum value, each $d$ value corresponds to an absolute error accumulated value, in which the minimum absolute error accumulated value is regarded as the best matching point, and the corresponding disparity is the obtained disparity value.
3.4. Experiment and analysis

3.4.1. Calibration experiment of binocular camera. Because pinhole model is only an approximation of the actual camera model, coupled with various lens distortions and deformations, the imaging model of the actual camera is much more complicated. Therefore, the distortion model is introduced, in which the distortion coefficient represents the distortion degree, so it is also a part of the camera internal reference. In this paper, Zhang Zhengyou calibration method is used for calibration, and 15 images of the left and right calibration plates (as shown in Figure 3) are collected. The calibration results of the left and right cameras are shown in Table 1 and Table 2.

![Figure 3 Calibration board image](image)

**Table 1** Internal parameters of left and right cameras

| Internal parameters | Left camera | Right camera |
|---------------------|-------------|--------------|
| Focal distance \( f \) /mm | 1403 | 1400 |
| Main point line value \( u_0 \) /pixel | 1125 | 1122 |
| Column point row value \( v_0 \) /pixel | 563 | 589 |
| Distortion coefficient \( k \) | -0.02 | -0.0011 |

**Table 2** Relative pose of left and right cameras

| External parameter | Translation vector T/mm | Spin matrix R/(°) |
|--------------------|--------------------------|-------------------|
| \( x \) axis       | -117.36                  | 0.044             |
| \( y \) axis       | 0.1024                   | -0.127            |
| \( z \) axis       | -1.033                   | 0.051             |

It can be seen from the data in the table that the actual size of the picture taken by the camera is 2108×1201, and the error with the calibrated data is very small; The actual horizontal distance between the two cameras is 120 mm, which is quite different from the calibrated data (117.36). The rotation matrix of the two cameras is close to the identity matrix, which indicates that the optical axes of the two cameras are approximately parallel.

3.4.2. Target grasping experiment. In the experimental system, firstly, the visual field image is shot by binocular camera, and then the stereo matching algorithm is used to detect whether there is a target object in the image. If there is a target, its position is determined by binocular vision. Finally, the computer transmits coordinate data to the robot, and the robot controller sends out corresponding control instructions to control the mechanical arm to move to the specified position and implement the grabbing action.

4. Methods of physical training for teenagers

4.1. Endurance training

Endurance running is a periodic high-intensity sport. With the help of physical training robot, it can cultivate lasting running ability and speed endurance. It requires athletes' central nervous system to have high functional stability, and cardiovascular function can adapt to sports function. From a
technical point of view, endurance running is characterized by running at a certain speed and lasting, with a large amount of exercise. It develops the ability to run for a long time by running over a long distance. From the physiological point of view, the performance of endurance running depends on the level of oxygen intake in the outside air and the working ability under anaerobic conditions, that is, aerobic metabolism and anaerobic metabolism.

4.2. **Strength training**

Strength quality is the foundation of human sports, the most basic quality of human sports and an important index to evaluate the level of physical training. Strength quality affects the mastery of sports skills, the improvement of sports achievements and the development of other qualities.

Strength quality training must also follow the principle of gradual and orderly progress, especially for young athletes and athletes at the beginning of a large training cycle or with long-term training interruption, the load must be gradually increased with the adaptation of sports support organs and the improvement of strength quality level, otherwise it will lead to the decline of training effect and even injury accidents. At the same time, we must take into account the characteristics of each athlete to carry out targeted training, adopt flexible and diverse training methods, and maximize the enthusiasm and initiative of athletes in training. Only in this way can we achieve the best training effect.

4.3. **Flexibility training**

Flexibility is one of the basic physical qualities, which is integrated with other physical qualities and coexists with each other. Physical qualities are very important for athletes. Physical fitness is the necessary guarantee for athletes to strengthen their physical strength, improve their quality, exercise their will and complete their training and competition tasks. It is also an auxiliary means to solve some specific technical problems. The better their physical fitness is, the longer their sports life will be. Because of poor physical fitness, other excellent athletes have to retire ahead of time. Because people's physical qualities are different, the speed of learning and mastering technology varies greatly. In order to continuously improve the technical level and meet the needs of competition, they must improve their physical qualities. Therefore, in attaching importance to the training of basic physical quality, special attention is paid to flexibility exercises.

4.4. **Matters needing attention in physical training**

To be able to train, we should start from a certain time, distance and quantity, and gradually increase the time, distance and the number of exercises and groups. At the same time, we should gradually improve the training intensity, and do some exercises close to the limit load, so as to give full play to the physical fitness of the human body.

It is not only the physical training, but also the cultivation and test of a person's will quality. Therefore, endure the unbearable pain of others in training, so as to cultivate their indomitable will and the spirit of never giving up until they reach their goals. The best way to develop muscle endurance is to increase the number of exercises and groups in each group under the condition of constant weight during strength exercises.

5. **Conclusion**

At present, Chinese scholars focus more on competitive sports, while foreign experts focus on mass fitness, and experts and scholars have not reached a consensus on some related issues, which shows that the study of young athletes needs more experts and scholars to explore constantly. In this paper, the target recognition and location grasping based on stereo matching algorithm are studied, and the automatic recognition and grasping action of the robot arm on the target object is realized. The experiment can identify and locate the target quickly and accurately in the small sample environment, which provides the premise for the robot arm to grasp the target. To provide theoretical basis and technical guidance for improving the level of physical training and sports performance of young athletes.
References

[1] Edelman S, Duvdevani-Bar S. A model of visual recognition and categorization. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, Vol. 352, no. 1358, pp. 1191-202, 2019.

[2] Zhang Xingxing. Analysis and Research on Physical Training of Youth in Aerial Skills of Freestyle Skiing. Sports Fashion, no. 1, pp. 32-32, 2020.

[3] Sun Yifei. Analysis of the principle of physical training for basketball teenagers. Boxing and Fighting, no. 6, pp. 61-61, 2020.

[4] He Liu. Talking about the strategy of physical training for table tennis teenagers. Chinese and Foreign Exchange, vol. 026, no. 024, pp. 76-77, 2019.

[5] Xing Yan. Discussion on the application of Cross Fit training mode in adolescent physical training. Boxing and Fighting, vol. 000, no. 006, pp. 103, 2019.

[6] Liu Q, Liu C. A Novel Locally Linear KNN Method With Applications to Visual Recognition. IEEE Transactions on Neural Networks & Learning Systems, vol. 28, no. 9, pp. 2010-2021, 2017.

[7] Donahue J, Hendricks L A, Rohrbach M, et al. Long-term Recurrent Convolutional Networks for Visual Recognition and Description[C]// IEEE Transactions on Pattern Analysis & Machine Intelligence. IEEE, 2017:677-691.

[8] Hou Ningjun, Zheng Li. Thoughts on youth physical training based on sunshine sports background. Track and Field, vol. 000, no. 009, pp. 18-19, 2017.

[9] Zheng Li, Hou Ningjun. Problems and Solutions of Youth Physical Training. Track and Field, no. 8, pp. 34-35, 2017.

[10] Lin T Y, Roychowdhury A, Maji S. Bilinear Convolutional Neural Networks for Fine-grained Visual Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017:1-1.

[11] Wu L, Wang Y, Li X, et al. Deep Attention-Based Spatially Recursive Networks for Fine-Grained Visual Recognition. IEEE Transactions on Cybernetics, 2018:1791-1802.