Structural design of anthropomorphic robot vision system

Junyan Zhang
Jiangsu Union Technical Institute Suzhou Industrial Park Branch, Suzhou, China

Corresponding author and e-mail: Junyan Zhang, 108645502@qq.com

Abstract. Anthropomorphic robot is one of the latest research directions in the field of intelligent robot research. It is an intelligent robot with similar shape and humanity. It has mobile function, operation function, sensing function, memory and autonomy, and can realize human-computer interaction. Robot technology has always been a hotspot of technology that people pay attention to. Because of the diverse development of robots, machine vision systems are gradually being established. The machine vision system realizes the functions of collecting images, analyzing images, acquiring information, processing images and others by simulating the human eye. It is a comprehensive technology integrating automation and computer. The article deeply studies the hardware composition and image processing technology of machine vision system and gives an example study on the application of machine vision in various fields. It also analyzes the current development status of machine vision system in China and looks forward to the bright future of machine vision.

1. Introduction
With the rapid development of science and technology, the development of robots has been nearly a hundred years old. In the medical field, medical robots assist or even replace humans to diagnose and operate patients. In the industrial sector, robots replace humans for inspection and packaging of products. In addition, robots have also been used in the military field, such as helping soldiers observe the battlefield and assist shooting. Among them, the machine vision system uses the digital camera to collect various kinds of information, analyzes and processes the image data through the central processor, and obtains useful information in the picture to help people complete the tedious and monotonous work. In the development of machine vision technology, it also keeps pace with the trend of social development. From the popular PS to the current face recognition technology, it gradually moves toward artificial intelligence, adding more convenient channels for people's lives. This article will be on the machine vision system. The hardware and software are described separately.

2. Design principles
Autonomous Mobile Robot (AMR) refers to an autonomous robotic system with athletic ability, which combines the characteristics of anthropomorphic robots and autonomous robots. Due to its wide and attractive application prospects, the research of autonomous humanoid robotics involves many subject areas such as robotic mechanical structure design, artificial intelligence, computer vision, motion control, planning decision-making, image processing, pattern recognition, information fusion, etc. There are a lot of challenges in the research, so the autonomous humanoid robot technology has become one of the focuses of the current research in the field of robotics. One of the key technologies of autonomous anthropomorphic robots is how to make them robust to autonomous navigation. Autonomous anthropomorphic robots can only have a purposeful movement when they know where they are and where the various targets are. At the same time, they can detect changes in the
environment and their own state in real time, and adjust their behavior accordingly to effectively accomplish specific tasks. Therefore, in order to achieve autonomy, the first task of anthropomorphic robots is to obtain an understanding of the working environment and its own state, that is, to realize the perception of the robot, which is one of the three basic elements recognized by the robot (i.e., perception, planning, execution).

Among them, the visual sensor can provide the most abundant sensory information compared with other robot navigation and positioning sensors. It is also a passive sensor, which is closest to the human environment perception mode, and its price is relatively moderate. Scientific research and statistics show that about 75% of the information that humans receive from the outside world comes from the visual system. Vision sensors also have the advantage of being able to work in high-rise urban environments or indoor environments, which is not possible with GPS receivers only; it can work with a large amount of highly dynamic occlusion, which is difficult to achieve based on laser, ultrasonic and other distance sensors; there is no accumulation of inertia sensor error with working time. At the same time, theoretical and technical advances in related fields such as computer vision and pattern recognition have also promoted the application of visual sensors in autonomous humanoid robots, such as target recognition, scene recognition, and self-positioning. The application of vision to autonomous humanoid robots has attracted the participation of a large number of researchers around the world, and has also achieved a lot of research results, such as IEEE Transaction on Robotics, IJCV, RAS, CVIU and other famous journals and ICRA, IROS, ICCV, etc. A large number of papers have been published at mainstream conferences such as CVPR and ECCV. Vision sensors have become an indispensable sensor for autonomous humanoid robots, and computer vision research and anthropomorphic robot research have also achieved good integration. Among various vision sensors, the omnidirectional vision device can obtain the panoramic information of the surrounding area of the robot in one image because of the horizontal angle of view of 360°. Through image processing, analysis and understanding, the target recognition and construction of the robot can be realized. Graphs, self-positioning, etc., have been used more and more in various anthropomorphic robots. This chapter will detail the history, development and application of omnidirectional vision in the following sections, especially in the application and challenges of autonomous anthropomorphic robots.

3. Hardware structure
The machine vision system gives the robot a view and can quickly analyze, resolve, and process information on complex images. The application based on this robot has been broadened, reducing the time for humans to handle cumbersome work and making our lives more convenient and efficient. In the pharmaceutical industry, the production process of drugs requires strict management control and quality control, and the machine vision system can realize this series of cumbersome operations and maintain the safety of patients. In laser processing, laser processing technology combined with machine vision technology makes processing more precise and reduces processing costs. The machine vision system consists of five parts: light source, lens, camera, image acquisition card and image processing system.

The light source is one of the important structures of the machine vision system. There are different requirements for the type of light source in different situations. Generally, it is a heat radiation power source, a gas discharge light source, a solid discharge light source, and a laser. LEDs are the most commonly used among light sources, and LEDs are divided into many clocks because of their different shapes. For example, when we perform microscope illumination, we use a ring light source to highlight the three-dimensional information of the object. Use a backlight to detect scratches on transparent objects, highlighting the contours of the object. Not limited to the LEDs mentioned above, the appropriate light source can be selected in different scenarios to capture valid picture information.

The lens is an important component of the image acquisition system of the machine vision system. It is like the human eyeball to observe the distance and range of the object under test. Its types are standard, telecentric, wide-angle, telephoto and close-up. People choose according to the object distance, shooting range, aperture, focal length and other conditions.

The camera is the core component of image acquisition. The light signal is converted into an electrical signal by a photosensitive element, and the conversion of light and digital signals is realized, and the original image is collected. Cameras are also divided into a variety of ways depending on the classification method. According to the chip classification, the most common types are CCD camera and CMOS camera. They are digital cameras for photoelectric conversion. However, due to different chip types, the processing power is also different. CMOS technology is more energy efficient than CCD. More efficient, but more expensive.

The image acquisition card is an indispensable part of the image processing part. In the processing of video information, the image acquisition technology is essential to digitize the image and store it in the form of a data file in the hard disk of the computer. Image acquisition generally relies on a large number of photosensitive
elements. For example, when the resolution is 1080×720, there are 1080 small rectangular rectangles on the photosensitive element and 720 small rectangles on the longitudinal side, waiting for the light intensity to be collected. Since the sensitivity of the light intensity received by the photosensitive elements is different, so that their resistance changes, the effects they produce will also differ to obtain the initial image. The larger the pixels that make up the picture, the clearer the image is, forming what people call high-definition pictures. In the machine vision system, the image processing system performs a series of analysis and processing through the image data obtained by the image acquisition card to obtain the desired information. First, the image information is collected by an image acquisition card, and the optical signal is converted into an electrical signal. The digital image is transmitted to the processor for grayscale processing, turning the part we are interested in into white, and the uninteresting part turning black, thus distinguishing between what we need and what we don't need. However, after obtaining the binarized image, there will be some noise around the target object, which affects the value of the edge of the subsequent judgment object. In order to eliminate these noises, we use image etching and expansion processing/filtering processing to corrode the target object, and then use the convolution operation to amplify the target, remove noise, and achieve boundary detection and image enhancement. After comparing with the database to obtain accurate information of the target in the field of view, based on this, the judgment and operation of subsequent execution.

4. Omnidirectional vision system design

Various designs of omnidirectional mirrors satisfying single-view imaging. The common feature of these mirrors is that the mirror shape is simple, the profile curve can be accurately described by mathematical analytic, and the processing difficulty is relatively small, and the omnidirectional configuration using these mirrors is relatively small. The main drawback of the vision system is that the obtained panoramic image has a large barrel distortion, and the resolution of the target imaging decreases rapidly as the distance from the visual system increases, and the imaging of the distant scene is small. A typical image captured by a hyperbolic vision system using a hyperbolic mirror in a 12m x 7m RoboCup medium group competition venue is shown in Figure 1. This omnidirectional vision system does not use football robots for a wide range of target recognition and accurate measurements for applications such as the RoboCup medium group robot soccer game. In response to the above-mentioned shortcomings of the single-view omnidirectional vision system, the researchers proposed a mirror design method that satisfies the specific imaging resolution requirements of various parts of the environment around the omnidirectional vision system, and also designed three imaging resolutions. The mirror with the same rate is the horizontal mirror, the vertical mirror and the angle mirror, which can realize the horizontal direction (perpendicular to the mirror rotation axis), the vertical direction (parallel to the mirror rotation axis) and the specular incident ray angle. The imaging resolution on the top is constant. None of these three mirrors can be used to construct an omnidirectional vision system that satisfies single-view imaging. This section first introduces the design of horizontal mirrors and vertical mirrors, and on this basis, designs a new combination mirror for the NuBot omnidirectional vision system.

Figure 1. Typical panoramic image captured by an omnidirectional vision system composed of a hyperbolic mirror

The horizontal equal-mirror mirror refers to a mirror surface that satisfies the distance relationship between the horizontal scene and the corresponding points in the formed image when the omnidirectional vision system is installed vertically. The mirror image is shown in Figure 2. F(t) is a function of the specular profile, and the point P(0, 0) on the horizontal plane to the optical axis with a distance d is imaged on the camera CCD far from the optical axis x. According to the principle of the angle of incidence, the angle of incidence is equal to the angle of exit, \( \phi = \psi + \theta \), so \( \tan(\phi + \theta) = \tan(\psi + 2\theta) \), and because \( \tan \theta = F(t) \), \( \tan \psi = x / f \), where f is the focal length of the camera, so there is
\[ d - \frac{1}{F} = \frac{\left( \frac{z}{f} + \frac{2F'}{1 - F'^2} \right)}{(1 - \frac{2F'}{1 - F'^2})} \]

\[ zF + 2F - z = 0, \text{ because the mirror is a convex mirror, take its positive root} \]

\[ F' = \frac{-1 + \sqrt{1 + \frac{z}{z}}}{z} \]

**Figure 2.** Schematic diagram of horizontal equal mirror design

5. **Calibration of the omnidirectional vision system**

The calibration of the vision system refers to determining the parameters of the mapping relationship between the two-dimensional image points on the imaging plane of the camera from the three-dimensional space point to the vision system. The vision system can only be applied to computer vision tasks such as visual perception measurement, environmental 3D reconstruction, visual system's own motion estimation, etc., and the accuracy of calibration directly determines the application effect. The omnidirectional vision system is no exception, and since the panoramic image obtained by the omnidirectional vision system has lower resolution and severe distortion compared to the ordinary fluoroscopic image, the calibration problem becomes more important. Since the omnidirectional vision system that satisfies single-view imaging has good characteristics, that each point in the panoramic image uniquely corresponds to an incident ray from the environment on the omnidirectional mirror surface, as shown in Figure 2.1, so based on imaging the omnidirectional visual calibration method of the model has been extensively studied and many achievements have been made. The existing calibration methods can be roughly divided into two categories: methods for prior knowledge of known environments, and methods that do not require the use of environmental prior knowledge. The first type of method requires three-dimensional coordinates of a series of spatial points in the environment in which the omnidirectional vision system is located, and the image points corresponding to these spatial points should be automatically extracted (or manually extracted), which we call These spatial points are control points. These control points are often obtained by setting a calibration block in the environment. The relationship between the three-dimensional space coordinates of the control points and the two-dimensional image coordinates of the corresponding image points can be used to recover the internal and external parameters in the imaging model of the omnidirectional vision system. Since the three-dimensional space coordinates and two-dimensional image coordinates of the control points are unknown, the calibration process is cumbersome. When there is no control point in the environment, this method requires several sets of lines parallel to each other in the known space, using the intersection of the conic curves imaged by these lines in omnidirectional vision (parallel lines in the panoramic image) the nature of the vanishing point is the internal parameter information of the omnidirectional vision system. The calibration idea is mainly based on the unified imaging model of various single-view omnidirectional vision systems proposed by Geyer and Danilidis. A detailed description of the model can be found in the literature. During the calibration process, the relative position between these lines and the relative position between the line and the omnidirectional vision system are not known. Ying and Hu derived a linear projection quadratic curve to provide three independent constraint equations for the internal parameters of the omnidirectional vision system, and a spherical quadratic projection curve to provide two independent constraint equations for the internal parameters of the system. Furthermore, a new calibration method for omnidirectional vision system based on geometric invariants is proposed, and it is concluded that the method based on spherical projection invariants can obtain more robust calibration results. Wu et al. also proposed a linear single-view omnidirectional vision system calibration method, which can simultaneously calibrate all omnidirectional vision systems except the main point by imaging.
information in at least three lines in a panoramic image. Internal parameters. The second type of method, also known as the self-calibration method, calibrates the omnidirectional vision system by using the polar line geometric constraint relationship between the pairs of points corresponding to multiple images. The polar geometry of various single-view omnidirectional vision systems is based on their respective imaging models. For a detailed derivation, see. The advantage of this method is that there is no need to use calibration blocks, and there is no need to know any prior information of the environment. The disadvantage is that it is necessary to be able to obtain corresponding pairs of points between images robustly and accurately.

6. Camera parameter automatic adjustment algorithm for image entropy

In digital cameras and home DVs, researchers have proposed some parameter adjustment mechanisms to improve imaging effects, such as changing the aperture or shutter time for automatic exposure, auto white balance, and auto focus. In some multi-slope response cameras, the researchers adjusted the response curve through automatic exposure control to adapt the camera's dynamic response range to different lighting conditions. But these methods are based on the camera hardware level, because we can't operate the internal hardware of most cameras (except some special hardware-supported cameras) of the robot vision system, so we can't use these methods. This paper attempts to adjust some parameters supported by the camera through external software to improve the imaging effect of the vision system under different lighting conditions. Some related research is currently in the RoboCup medium group. The RoboCup Medium Group Robot Soccer is a standard test platform for robot vision related issues. Although the latest rules of the game have undergone some major changes, such as turning the yellow/blue goal into a net like a human race, the colored columns have been cancelled, but the game environment is still largely color-coded. The ultimate goal of RoboCup is that the robot soccer team can defeat the human world champion. The robot will need to be able to compete in a highly dynamic environment or even an outdoor environment, so how to make the designed vision system robustly recognize the color coding. The goal remains a research focus for RoboCup researchers. In addition to adaptive color segmentation methods, color online learning algorithms, and color recognition-free target recognition algorithms, some researchers have also attempted to implement visual sensors by adjusting camera parameters. Robustness. Reference defines the camera parameter adjustment problem as an optimization problem and uses genetic algorithms to minimize the distance between the actual color value of the pixel and the theoretical color value in the artificially selected image region to obtain the optimal solution to the problem. Since the theoretical color value is used as the reference value, the influence from the ray condition can be eliminated, but the method requires manual selection of some special image regions as reference information. Reference designed a set of PID controllers that adjust camera parameters such as gain, aperture, and two white balance channels based on the pixel color values of the white areas that are always visible in the imaging of the omnidirectional vision system. Reference designed a PI controller to adjust the exposure time to adjust the color value of the reference green area to the desired color value. Reference proposes to automatically set camera parameters such as exposure time, gain and white balance of the omnidirectional vision system based on the luminance information of the entire image and the information of the known black and white regions in the image, but the method requires Each black and white swatch is placed in advance on the field, so it can only be used for offline adjustment calibration before the game. These methods require some reference colors in the camera parameter adjustment process, thus limiting their application in many other applications.

Using an omnidirectional vision system to obtain a series of panoramic images at different exposure times and gains in indoor and outdoor environments and calculating the image entropy of all images according to equation, it is possible to analyze how the image entropy changes with camera parameters. The indoor environment is a standard 18m x 12m RoboCup medium-sized group competition venue. The lighting conditions are not only dependent on the indoor lighting, but also may be greatly affected by natural light through a large number of glass windows. The outdoor environment contains two blue swatches and some elements of the indoor environment, such as a green carpet, two orange football balls and black obstacles. The experiments for all omnidirectional vision systems in this chapter will be performed in both environments. Since the lighting conditions of the two environments are completely different, and the dynamic response range of the cameras used is limited, two sets of omnidirectional vision systems with different aperture settings (the lens aperture can only be manually adjusted) are used in indoor and outdoor environments. (i.e. two soccer robots) to conduct experiments. In indoor environment experiments, the exposure time ranges from 5ms to 40ms, and the gain ranges from 5 to 22. The time of the experiment is night, so the lighting conditions are only illumination, not affected by natural light. In outdoor environment experiments, the exposure time ranges from 1ms to 22ms, and the gain ranges from 1 to 22. The weather is cloudy, and the experiment time is noon. The minimum adjustment steps for the two parameters are 1ms and 1 respectively. During the experiment, a panoramic image is captured after each set of camera parameters is set.
Considering that all images corresponding to the image entropy on the ridge curve are suitable for robot vision, by defining a certain search path, the two-dimensional optimization problem can be converted into a one-dimensional optimization problem. Since the RoboCup mid-group competition is a highly dynamic and color-coded environment, high exposure times and low gains can cause real-time performance degradation of the system, while low exposure times and high gains can cause more noise in the image, which in turn causes color segmentation and other image processing results to deteriorate. Therefore, it is necessary to make a compromise between the exposure time and the gain, and neither of the two values should be too high. Depending on the parameter range of the camera used, the search path is defined as "exposure time = gain" (since the exposure time is in milliseconds, the gain is unitless, so it is only equal in value), and the maximum image entropy is searched along this path. The camera parameters corresponding to the image entropy are the optimal parameters under the current environment and current lighting conditions. According to the imaging characteristics of omnidirectional vision, the robot itself is imaged in the central portion of the panoramic image. Therefore, in practical applications, the robot can determine whether the robot has entered a new working environment or whether the lighting conditions in the current environment have changed by calculating the brightness average of the central region of the image. If the increase in the mean exceeds a certain threshold, the robot assumes that the ray condition becomes more intense, so the optimization of the camera parameters will follow the search path and toward the exposure time and gain increase. Similarly, if the decrease in the mean exceeds the threshold, the optimization process proceeds along the search path and toward the exposure time and gain reduction. In the experiment, the threshold was set to 20. During the optimization process, the mean value of the brightness in the center of the image does not need to be recalculated, and the change does not need to be considered. After the optimal camera parameters are obtained, the brightness average is recalculated and saved, and compared with the new value to determine whether adjustment and optimization of the camera parameters are needed. Since the lighting conditions do not change all the time, this judgment only needs to be performed once every certain time.

When the robot is located at different positions in the field, the content contained in the image is different, and the maximum image entropy on the search path changes as the position of the robot changes. Therefore, once the robot finds that the camera parameters need to be adjusted, it will keep on the field until the camera parameter adjustment process is completed. Moreover, according to the discussion in the following section 3.4, the camera parameter adjustment can be completed in a few hundred milliseconds. In the optimization process, the surrounding environment of the robot can be approximated as static, and the image entropy only changes with the camera parameters. The change, so the maximum image entropy and the optimal camera parameters can be easily searched, and the optimization of the camera parameters is not affected by the surrounding environment of the robot. This paper will also verify the camera parameters optimized in the static environment through experiments. It is also an ideal parameter in a highly dynamic environment. During the optimization process, a new set of camera parameters is set to the camera, and the robot acquires a new panoramic image and calculates its image entropy according to equation. The new image entropy is used to compare with the image entropy obtained in the previous optimization step to determine if the maximum value has been reached. This iterative process continues until the maximum image entropy is searched. On how to choose new parameters, the algorithm uses variable step size techniques to speed up the search process. When the current image entropy distance Max E is not large, the optimization step size is set to 1, that is, the change of exposure time is 1 ms, the change of gain is 1; when the current image entropy distance ( )cMax E is large, the optimization step size can be set to 2 or 3. Depending on the camera characteristics and the different requirements of the vision system for different applications, the camera parameters need to be adjusted and optimized along different search paths. For example, different cameras have widely different gain ranges. Some cameras have gains from 0 to 50, and some have gain ranges from 0 to 4000. Therefore, the search path can be defined as "exposure time = α * gain" (which is also only numerically equal), and the parameter α can be determined after analyzing the influence of the camera gain on the image. If the camera's parameter range is similar to that of the camera used in this paper, in some cases, the signal-to-noise ratio of the image is required to be high, but the real-time performance of the vision system is not too high. The search path can be defined as "exposure time = α * gain". Where α > 1, if other occasions require the camera to output the image as fast as possible, and the image noise requirements are not strict, then the search path can be defined as "exposure time = α * gain", where α < 1.

7. Machine vision applications

In the military field, machine vision systems play a very important role in aviation, aerospace, and tracking. The Air Force mainly uses drones to complete air missions, such as targeting and accurate annihilation. In the case of high altitude and movement, by matching the infrared sensor, the picture information is returned in real time, and the target object of the database is matched. The similarity is extremely high, and a strong infrared signal is detected, and the enemy information such as illegal nuclear weapons is judged in the area, to accurately strike the target object. Reduce personnel losses and improve operational accuracy. When performing the high-altitude
reconnaissance mission of the drone, the high-resolution image acquisition and analysis processing helps us to monitor the surrounding environment at any time and prevent the enemy from invading while observing the daily operation of the enemy.

In the industrial field, product quality inspection of repetitive machinery consumes a lot of labor costs, and the processing speed is slow, which is affected by disturbance factors such as weather, physical state and working time. The robot vision system is not affected by various interference factors and has an excellent processing system to quickly identify the quality defects of large-volume products in a few seconds, which is fast and accurate. It is calculated that the industrial robot vision sorting system can complete the workload of 500 workers a day, greatly improving the efficiency of production and liberating the labor force. Industrial vision robots can continuously learn, and each robot can quickly refresh knowledge through the network, continuously learn and consolidate, reduce error links and improve accuracy, which is equivalent to having numerous reproducible senior inspectors. Not only in the inspection work, the factory intelligent robot can also use the machine vision positioning to quickly locate the machine on the assembly line, quickly analyze the material and perform processing operations.

In the civilian field, the scope of application of machine vision recognition technology is very extensive, and there are deep contacts in face verification, security, and transportation. We can see it on the computer in the security room. The computer obtains the license plate number of the vehicle entering and leaving the vehicle through the captured picture and compares whether the vehicle is the license plate number of the community, and calculates how long the vehicle enters and exits the area. It simplifies the vehicle entry and exit process to a certain extent, providing a safe and fast access for vehicles to enter and exit. When entering the airport check-in, the automatic ticket checking machine compares the face with the ID card photo, quickly and accurately judges, saves labor costs, and speeds up airport security. In addition, the camera is used to detect suspects in real time, and the photos taken by various cameras are compared with the photos of the prisoners in the database to analyze the similarities, so that criminals can be identified and arrested in the vast sea of people.

8. Conclusion
The omnidirectional vision system has become one of the more important sensors for autonomous anthropomorphic robots and has extensive attention and research on it. This paper focuses on how to improve the robustness of omnidirectional vision system and its application in target recognition and self-positioning. According to the software system of the anthropomorphic robot vision system designed in this paper, the image captured by the dual camera is better realized, and the modular structure with easy expansion is provided for the realization of algorithms such as image processing, ranging and recognition in the future, better software and hardware platform.

References
[1] Liu Huacheng. Design of Robot Vision System [J]. Science and Fortune, 2013(12):11-12.
[2] Du Junxi, Huang Xinhan. Design of Chess Robot Vision System [J]. Electronic Technology Application, 2007, 33(9): 133-136.
[3] Guan Bing, Gu Jinan, Chen Sijie, et al. Vision System Design of Cylinder Workpiece Loading and Unloading Robot [J]. Mechanical Design and Research, 2012, 28(1): 18-21.
[4] Zhao Zhengang, Lü Jianping. Design of Monocular Infrared Target Tracking Robot Vision System [J]. Journal of Optoelectronic Technology, 2008, 23(6): 19-23.
[5] Shen Yige, Qi Guo. Design of Picking Robot Vision System Based on DSP and μC/OS-II [J]. Journal of Agricultural Mechanization Research, 2017(06): 229-233.
[6] Li Lei. Research on anthropomorphic robot system design and visual navigation control [J]. Graduate, 2003.
[7] He Xuejian, Dong Guanqiang. Design of Monocular Vision System for Wheeled Anthropomorphic Robot [J]. Mining Machinery, 2005(02): 77-78.
[8] Qi Di, Li Xin. Design of Soccer Robot Vision System Based on Real-time Multi-target Tracking [J]. Mechanical & Electrical Engineering, 2001(06): 40-42.
[9] Zhu Xi, Li Yongxin, Chen Sheng, et al. Design of Soccer Robot Vision System Based on Multi-target Real-time Search--Robocup Series Research [J]. Automation and Instrumentation, 2004(2).