EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Mobile Robot Vision Image Feature Recognition Method Based on Machine Vision

Qin Dong

School of Information Engineering, Yancheng Institute of Technology, Yancheng, 224000, China

Abstract

The rapid development of machine vision and the widespread application of mobile robots in various environments have posed new demands and challenges for efficient visual image feature recognition. To improve the efficiency and accuracy of mobile robot visual image feature recognition, a mobile robot visual image feature recognition method based on machine vision is proposed in this paper. Firstly, the development of mobile robot vision is analyzed, and the specific functions of robot visual feature recognition method are designed. Then, the Fourier series method is used to collect the mobile robot visual image, and the matrix associated with the auto-correlation function is calculated according to the Harris algorithm to complete the edge feature extraction of the mobile robot visual image. SIFT feature points of mobile robot visual image are classified, and mobile robot visual image feature recognition is realized through machine vision. The experimental results showed that when the number of images was 600, the accuracy of image feature recognition and the loss value of image edge feature extraction of this method were 96.98% and 6.38%, respectively, and the number of iterations was 500. The time of visual image feature recognition of this method was only 3 minutes. The method had the lowest error mean and error variance under different noise conditions. This method can effectively improve the efficiency and accuracy of image feature recognition, promote the development of machine vision and mobile robot technology, and stimulate new research and applications.

Keywords: SIFT feature point classification; Fourier series; Harris algorithm; Visual image

Received on 14 June 2023, accepted on 15 December 2023, published on 20 December 2023

doi: 10.4108/ew.3450

[1]Corresponding author. Email: dong_qin22@163.com

1. Introduction

In today's era of rapid development of science and technology, a variety of advanced scientific and technological products with various functions have entered thousands of households, which was unimaginable before, such as mobile terminal equipment. Since the mid-1980s, robots have transitioned from the structured environment of factories to the daily living environment including shopping malls, restaurants, and households. This expansion has penetrated further into chaotic and uncontrollable environments. Intelligent service robots have advanced significantly in recent years. They can complete tasks independently, collaborate with humans, or carry out tasks under human guidance. The progress of mobile robots acting as shopping assistants, caregivers, and receptionists is particularly noteworthy [1]. Robot technology has become a new generation of revolutionary technology after computer technology, which will affect the development pace of the whole society. Robots can be broadly divided into industrial robots and service robots, which has different application environments. Industrial robots, such as drying and code robots, are commonly utilized. On the other hand, service robots like restaurant service and humanoid robots are often employed in indoor settings [2]. After years of collection and sorting, the International Federation of Robots has given a preliminary definition to the service robot: the service robot is a semi-autonomous or fully autonomous robot, which can complete the service work intended for human health but does not include the equipment engaged in production.

An important sensing source for mobile robots is the visual sensor. However, early research on this topic was often abandoned by many researchers due to two major defects: high hardware costs and lengthy computing times [3]. Due to the emergence of large-scale integrated circuits and the improvement of computing speed, visual sensor performance, and price decline, modern visual mobile robots have developed rapidly. In addition, the continuous changes in image processing and visual technology, as well as the objective development prospects and importance of visual mobile robots in military applications, have also promoted its rapid progress. The robot competitions held around the world and some influential international robot competitions have all promoted the development of this website [4]. At first, the vision system of mobile robot was only applied to some specific occasions. Later, with the continuous strength of its function, it developed to simulate the function of human eyes. Now, the vision system has become increasingly practical. Summarizing its development process, it can be divided into four distinct stages:

The initial stage of the paper is from the 1960s to 1970s. At this time, robot vision was only academic, but its architecture and image processing process were studied theoretically. At that time, Stanford Institute, French National Research Center and other institutions made outstanding contributions in this field [5]. From the end of the 20th century to the mid-1980s, robot vision developed to the military stage. At this time, the main purpose of its research was to design and develop a series of vehicles that can move autonomously in an unstructured environment for the military, including vehicles and transportation robots. The Autonomous Landed Vehicle (ALV) project at that time was a very favorable description, and the high-speed km intelligent vehicle technology derived from the project made outstanding contributions to the civilian application of robot vision [6]. The subsequent phase entailed supplementary scientific research that focused primarily on Mars rover and autonomous vehicles utilized in related scientific investigations. This research also achieved various technological advancements and innovations. Currently, the practical stage of robot vision development has been reached. Forgotten research findings, in conjunction with rapid advances in software and hardware technology, have facilitated the gradual integration of visual mobile robots into daily life. As a result, their functions have become more practical and increasingly intertwined with people’s lives.

Despite undergoing several stages of rapid development, the vision of mobile robots remains significantly lower compared to human vision abilities. However, advancements in large-scale integrated circuits, machine vision, and artificial intelligence make it possible for mobile robots to eventually have functions comparable to human eyes. Therefore, relevant scholars have made some progress in comparative research.

Kong Yan et al. proposed a human-behavior recognition method based on visual attention [7]. Using the depth convolution neural network of visual attention, they added a weight to the video image features to pay visual attention to the beneficial areas in the features. Experiments were carried out on the self-built oilfield-7 oilfield data set and hmdb51 data set to verify the effectiveness of the proposed network model suitable for human behavior in oil field. This method could improve the effect of human behavior recognition. Zeng Jinle et al. proposed an automatic recognition method for weld trajectory based on multi visual feature acquisition and fusion [8]. It combined multiple visual information of the weld seam area for comprehensive decision-making, fully utilizing the redundancy and complementarity between different visual feature information to accurately identify the position of the weld seam trajectory. Thus, the deviation between the actual welding trajectory and the machine teaching trajectory was compensated in real-time, improving the accuracy of welding seam trajectory recognition. Xue Teng et al. presented a technique for stable robot gripping that relies on visual perception and prior tactile knowledge learning [9]. The authors assessed the gripping performance by measuring the object's resistance to external disturbances during the gripping process. On this basis, the visual tactile joint data set was established, and the tactile prior knowledge was learned. The stable grasp structure was formed through the fusion of visual and tactile data in the robotic grasp system. Ten target objects were experimentally verified. The stability of the grasping method had been improved resulting in a good robotic grasping effect, although the efficiency of stable grasping remained low.

Improving the visual recognition ability of mobile robots is of great significance for enhancing their intelligence, safety, and accuracy, especially in automation and interactive tasks. This article proposes a machine vision-based image feature recognition method for mobile robots to develop advanced image processing algorithms, achieve environmental awareness, and enhance the autonomous decision-making ability of mobile robots.

2. Robot Visual Feature Recognition Method

2.1 Method and Function Design of Robot Visual Feature Recognition

Robot vision technology aims to create a vision system for robots that enables them to perceive the environment as flexibly and intelligently as human vision system and make corresponding processing in time. Bottom vision, middle vision and high vision are three different levels of vision technology, as shown in Figure 1 [10].

Robot vision is a technology that enables automatic image-based detection, control, and analysis. In robot vision system, computer is used to simulate human visual objects. In establishing a visual information system for computer-assisted human completion of visual tasks, application of image understanding and recognition in photographic geometry, probability theory, random processes, artificial intelligence, and related theories are necessary [11]. For example, human eye recognition and robot vision need the help of two kinds of intelligent activities: perception and thinking.

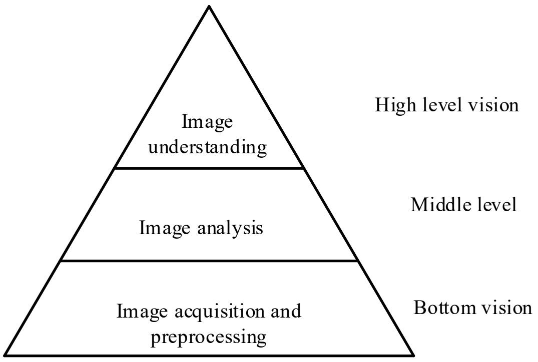

Fig.1. Schematic diagram of robot vision technology level

The robot visual feature recognition method consists of two parts: hardware and software. The hardware part can be regarded as the skeleton and body of the robot vision system, including image acquisition components (such as Charge Coupled Device image sensor or Complementary Metal Oxide Semiconductor camera), video signal digital conversion components (such as image acquisition card) and video signal central processing components, as well as processors (such as Digital Signal Processor based fast processor, single chip microcomputer and systolic structure) [11]. There are generally two ways of image acquisition: monocular vision and stereo vision. Monocular vision is a vision system with one vision sensor. Stereo vision generally refers to a vision system with two vision sensors. Monocular vision has the advantages of simple structure, short measurement time and low program complexity. However, for applications that demand high accuracy, monocular vision exhibits limited robustness. Binocular vision can make up for the deficiency of monocular vision in the case of high accuracy requirements [12].

The software part is the soul and idea of the robot vision system, including the development platform of the software system (computer software), the realization of the software, the functional algorithm and the robot control software. This part is mainly the implementation of image processing theory and algorithm.

The composition of robot visual feature extraction is shown in Figure 2.

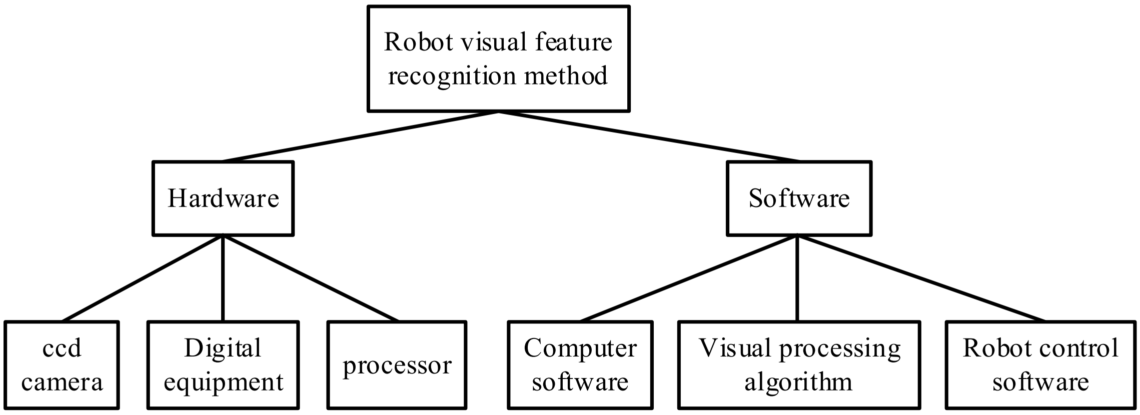

Figure 2: Design of robot visual feature recognition function

2.2 Mobile robot vision image acquisition based on Fourier series

Sampling theorem can be divided into the one-dimensional sampling theorem and the two-dimensional sampling theorem according to different dimensions.

The one-dimensional function $g(t)$ is defined as a time continuous analog signal, and it is represented by sampling sample $g(kt)$. k -a represents the whole value. T is the sampling period [13].

The method of reconstructing the original function $g(t)$ from the sample $g(kt)$ is to interpolate at the appropriate place among the samples. Generally, the following interpolation function $g(t)$ can be used:

$$g(t)=\sum_{k=-\infty}^{\infty} g(kT)h(t-kT) \qquad (1)$$

$h(t)$ is the time that interpolation function kT moving along the t axis. The effect of sample $g(kt)$ on function $g(t)$ at time t is weighted by the coefficient $g(t-kT)$. Now it is assumed that both g and h can perform Fourier transform:

$$f(kT)g(t-kT)=\sum_{k=-\infty}^{\infty} f(\tau)g(t-\tau)\delta(t-kT) \quad (2)$$

the

$$f(t)=\int_{-\infty}^{+\infty} f(\tau)g(t-\tau)\sum_{k=-\infty}^{+\infty}\delta(\tau-kT)d\tau \quad (3)$$

In formula (3), τ and δ are both parameters. $\sum_{k=-\infty}^{+\infty}\delta(\tau-kT)$ is a periodic function with period T, which can be expanded into Fourier series.

$$\int_{-\infty}^{+\infty}\delta(\tau-kT)=\sum_{-\infty}^{+\infty}a_n\exp(j2\pi nt/T) \quad (4)$$

The Fourier expansion coefficient a_n can be obtained from the following formula (5):

$$\begin{aligned} a_n &= \frac{1}{T}\int_{-T/2}^{T/2}[\sum_{k=-\infty}^{+\infty}\delta(\tau-kT)]exp(\frac{-j2\pi nt}{T})d\tau \\ &= \frac{1}{T}\int_{-T/2}^{T/2}\delta(\tau)exp(\frac{-j2\pi nt}{T})d\tau \end{aligned} \quad (5)$$

In the integral above, the only time a_n is not equal to 0 when $k=0$.

$$a_n=1/T \quad (6)$$

As shown in formula (7), $g(t)$ can be expressed as the sum of the convolution of f and $\frac{h(t)}{t}$ of $g(t)exp(\frac{j2\pi nt}{T})$.

$$g(t)=\sum_{n=-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(\tau)exp(\frac{j2\pi nt}{T})+\frac{h(t-\tau)}{T}d\tau \quad (7)$$

According to the convolution characteristics of Fourier transform, it can be concluded that the transformation of each term in the summation sign in formula (7) is the product of the Fourier transforms of two functions [14]. $G(\omega)$ and $H(\omega)$ represent the Fourier transform of $g(t)$ and $h(t)$, respectively. That is to say, Fourier series can be used to collect visual images of mobile robots.

The robot vision system primarily focuses on enabling robots to emulate the human and organism's visual feature recognition function. This enables it to perceive, conceptualize, and evaluate its surrounding environment, thereby achieving its recognition and comprehension objective. The primary objectives of recognizing image features through robot vision include the acquisition of images, preprocessing, image segmentation, description of features, recognition and classification, comprehension of three-dimensional information, depiction of scenes, image interpretation, and more, as indicated in Figure 3.

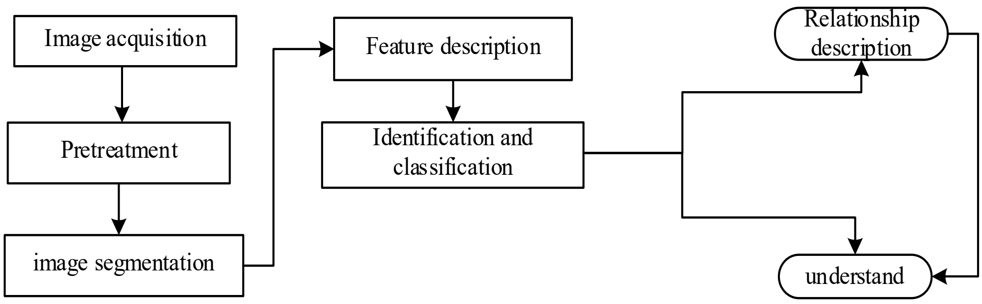

Figure 3: Visual task flow chart of mobile robot

Based on the research of two-dimensional image recognition algorithms, this paper proposed a real-time point cloud image recognition algorithm, which is a recognition and judgment method that integrates feature space and the minimum distance of the same element. The effectiveness of the real-time recognition method was verified on the computer, and successfully integrated into the robot system. The experimental data were analysed and processed. Finally, the problems in the current robot vision system were analysed, and some suggestions were proposed for the design of the next generation robot vision system to optimize the robot vision system.

3. Mobile Robot Visual Image Feature Recognition Based on Machine Vision

Chapter 2 mainly introduces the methods of robot visual feature recognition, explains the methods and functional design of robot visual feature recognition, and mobile robot visual image acquisition based on Fourier series. On this basis, Chapter 3 will conduct research on mobile robot visual image feature recognition based on machine vision, including edge feature extraction, edge feature extraction, and classification SIFT feature point acquisition of mobile robot visual images, in order to better improve the research on mobile robot visual image feature recognition.

Image feature extraction is a key problem in the field of computer vision image processing. Image feature extraction exists due to machine vision. To recognize the image, the computer extracts the relevant pixels composed of the image, and analyzes the pixels to determine their feature attribution, which is image feature extraction [15]. From the starting point of the first mock exam, it is a method to transform a set of measured values of a pattern to highlight the typical characteristics of the pattern. It can be used to identify the feature points in some regions as the input of continuous identification through image analysis and transformation. The starting point of subsequent processing is the image features. As the "interesting" part of image description, image features reflect the most basic attributes of the image itself, which can be quantified in combination with vision [16]. Image visual features are the description of image regions containing significant structural information of the image, such as edges, corners, and other image features. To detect the region of interest of an image, a salient feature measure is defined and calculated by the extreme values of image pixels and local regions. The purpose of examining different image sizes is to enable identification of the same image region, even if it exists within distinct scale spaces of various images. This process is called scale invariant detection. The extreme value of salient feature measurement is selected to ensure the repeatability of the inspection process. The definition of feature repeatability is because the same feature point may be detected in the same scene of two or more images [17]. In fact, there are many kinds of image features that can be extracted from digital images, including corner features, edge features and speckle features.

3.1 Corner Feature Extraction of Mobile Robot Vision Image

Generally speaking, a point is defined as the intersection of two edges. In a digital image, a point refers to the maximum value of the adaptive correlation function corresponding to the point pixel. In recent years, a series of point feature extraction algorithms have emerged in the field of image processing, which is mainly divided into two categories. The first kind of algorithm first extracts the image edge information, and then looks for the point with the maximum curvature value, or the intersection of edge segments as point features. The second kind of algorithm is mainly aimed at finding point features in gray image. Point features are defined as a point with two dominant directions and different edge directions in the local neighborhood of this point. The ability to detect the same point of the same image under different backgrounds, including varying lighting conditions, is a reflection of the quality of extracting point features [18].

Harris algorithm calculates the matrix associated with the autocorrelation function and sets the first-order curvature of the auto-correlation function as the eigenvalue of the matrix. When the row column curvature value of a point in the image reaches the maximum, the point is defined as the image point feature. The mathematical expression of Harris algorithm is as follows.

$$L = H(\tilde{s}) \otimes \begin{bmatrix} h_x^2 & h_x h_y \\ h_x h_y & h_y^2 \end{bmatrix} \tag{8}$$

h_x is the gradient in the x direction. h_y is the gradient in the y direction. $H(\tilde{s})$ is the Gaussian template.

The angular response function of Harris algorithm is:

$$P = \det(L) / ktr(L)^2 \tag{9}$$

$\det(L)$ is the determinant of the matrix. $ktr(L)$ is the direct trace of the matrix, and k is the default constant. The angular response criterion P is positive in the angular region, negative in the edge region, and small in the unchanged region. To judge whether the point is a corner by calculating the P value of the center point of the image window. If P is greater than a given threshold value, this point is considered as a corner [19]. Harris points feature extraction algorithm has the characteristics of simple calculation, uniform and reasonable corner features, quantitative extraction of feature points and stable operator. The feature points extracted by Harris algorithm are the pixels corresponding to the value of great interest in the local range. The threshold in Harris algorithm depends on the attributes of the actual image, such as size and texture. It does not have intuitive physical meaning, and the specific value is difficult to determine.

3.2 Edge Feature Extraction of Mobile Robot Vision Image

Line features include edges and lines. The meaning of edges is to distinguish local areas with different features, while lines are edge pairs that delimit the same feature area. Edge is very important for people to distinguish objects. Edge extraction is a basic and important problem in image analysis. In digital images, edges represent object boundaries. These distinct boundaries can aid people in directly identifying objects on many occasions. Therefore, edge feature extraction has important application value in the fields of image segmentation, image reconstruction and target recognition [20]. The edge is located where the brightness value of the two-dimensional image function changes suddenly and violently from one shape to another, such as from a white square area to a black background area. The edge is a collection of points, and these points are the extreme values of the local region of the image gradient.

For an image $I(x,y)$, x and y are the abscissa and vertical coordinates of a pixel respectively. The directional derivatives are h_x and h_y, respectively [21]. Based on the characteristics of gradient and direction distribution of pixels in the neighborhood of feature points, the gradient amplitude can be obtained as follows:

$$m(x,y)=\sqrt{(h_x)^2+(h_y)^2} \tag{10}$$

The direction of the gradient is formula (11).

$$\theta(x,y)=\tan^{-1}({}^{h_x}\!/\!_{h_y}) \tag{11}$$

In formula (11), $h_x=\dfrac{\partial I(x,y)}{\partial x}$ and $h_y=\dfrac{\partial I(x,y)}{\partial y}$.

The local maximum of the gradient amplitude of image $I(x,y)$ is found to be the basic idea of constructing the first derivative edge detector. An odd symmetric filter can approximate the first derivative, and the convolution output peak corresponds to the edge in the image. Usually, the first digital image derivative is expressed through the convolution of a digital image convolution template, referred to as an edge operator, and then the resulting output is processed to obtain a mapping of gradients. The value of the gradient mapping is calculated as the input of the non-maximum suppression process, and the local maximum of the mapping is finally set as a threshold to reduce the edge mapping. When the maximum value of the first derivative of the digital image is obtained and the second derivative of the digital image is zero, the zero-crossing point is found in the second derivative of the image $I(x,y)$ gradient to detect the image edge [22]. The typical zero crossing detection operator is Laplace operator:

$$\nabla I=\frac{\partial^2 I(x,y)}{\partial x^2}+\frac{\partial^2 I(x,y)}{\partial y^2}=h_{xx}+h_{yy} \tag{12}$$

The Laplacian operator is sensitive to noise, resulting in bilateral effects and an inability to detect the edge direction. It is generally not employed directly for edge detection due to these limitations.

In the field of computer vision, the main idea of speckle detection is to detect the region in the image that is larger or smaller than the surrounding pixel gray value. Typical speckle detection algorithms are divided into two categories: derivative based differential method, which is called differential detector, and watershed algorithm based on local extremum.

Detecting image spots using Gaussian Laplacian is the most typical spot detection method. The two-dimensional Gaussian function is defined as:

$$H(x,y,\sigma)=\frac{1}{2\pi\sigma^2}\text{-}\,exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{13}$$

Its Laplace transform is defined as:

$$\nabla^2 h=\frac{\partial^2 h}{\partial x^2}+\frac{\partial^2 h}{\partial y^2} \tag{14}$$

The normalized Gaussian Laplace transform is:

$$\nabla^2_{norm}=\frac{1}{2\pi\sigma^2}[1\text{-}\frac{x^2+y^2}{\sigma^2}]\text{-}\,exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{15}$$

The normalized algorithm is a circular symmetric function displayed on the two-dimensional image. This operator is used to detect spots in the image, and two-dimensional spots of different sizes can be detected by changing the value.

3.3 Mobile Robot Vision Image Classification SIFT Feature Point Acquisition

Scale Invariant Feature Transform (SIFT) is a common feature point extraction and description algorithm in computer vision. This algorithm can detect feature points with scale invariance and rotation invariance in images, which can be used for tasks such as image matching, localization, and recognition. The core idea of the SIFT algorithm is to detect stable feature points in images at different scales and directions. It detects image features at different scales by constructing Gaussian and differential pyramids. Then, the position of key points is determined by detecting local extremum points at each scale, and the scale space extremum suppression method is used to eliminate unstable edge responses. After detecting the position of key points, the SIFT algorithm calculates the main direction of each key point and describes the key points as feature vectors with rotation invariance. These vectors have good distinguishability and robustness, allowing for image matching and comparison that is not affected by image scaling, rotation, brightness changes, and other disturbances. The SIFT algorithm is widely used in the field of computer vision, especially in tasks such as target recognition, image stitching, 3D reconstruction, and object tracking. Its stability and robustness make it suitable for processing images with various perspectives, lighting conditions, and scale changes, which is why it is popularly used in image processing and computer vision applications.

In the SIFT feature point extraction stage, firstly, the scale space is established, and then the extreme points are found from the scale space. In Lowe's algorithm, the intermediate detection point corresponds to 8 adjacent points on the same scale and 9 adjacent points on the upper and lower scales. These two points are compared to 26 points to ensure that extreme points are detected in both scale space and two-dimensional image space. If a point is the maximum or minimum value in the 26 fields of this layer and the upper and lower layers of the dog scale space, it is considered as a feature point of the image under this scale.

Based on the 200 robot laboratory images obtained through the mobile robot vision system, the maximum eigenvalue points extracted from an image are almost identical to the minimum eigenvalue points. Additionally, all properly matched feature points originate from the same class of sift extreme points. Therefore, the extracted SIFT feature points are divided into two groups. In the feature matching stage, only the feature points belonging to the same type are compared. In this way, the matching speed is effectively improved without losing the correct matching feature points.

To calculate the feature matching time after classifying SIFT feature points, it is assumed that the number of features extracted from the two images are:

$$M = m_{\max} + m_{\min} \tag{16}$$

$$U = u_{\max} + u_{\min} \tag{17}$$

$m_{\max}$ and $u_{\max}$ are maximum eigenvalue points. $m_{\min}$ and $u_{\min}$ are minimum eigenvalue points.

It is supposed that the matching time of the original SIFT algorithm is:

$$R = m^*u \tag{18}$$

In feature matching, when only SIFT feature points of the same type are compared, the feature matching time is:

$$R_s = m_{\max}{}^*u_{\max} + m_{\min} * u_{\min} \tag{19}$$

Because the number of extracted maximums SIFT feature points is basically the same as that of minimum SIFT feature points in the same image, the formulas are obtained:

$$m_{\max} \cong m_{\min} \cong m/2 \tag{20}$$

$$u_{\max} \cong u_{\min} \cong u/2 \tag{21}$$

Through formula (20) and formula (21), the following (22) is obtained:

$$R_s = \frac{m * u}{2} = \frac{R}{2} \tag{22}$$

It is proved that the matching time of classification SIFT feature point matching method is reduced by 50% compared with the original SIFT algorithm. The robot laboratory images collected by the rehabilitation robot vision system are selected. The original SIFT algorithm and the classified SIFT feature point method are applied to carry out the feature matching experiment respectively. Some experimental results are shown in Table 1.

Table 1. Feature matching experiments of original sift algorithm and classified sift feature point method

Number of SIFT feature points extracted		Original SIFT algorithm		Classify SIFT feature points	
First picture	Second picture	Matching time(s)	Correct matching quantity(number)	Matching time(s)	Correct matching quantity(number)
645	732	0.686	237	0.311	311
777	640	0.790	264	0.330	395
676	621	0.760	205	0.306	383
671	621	0.810	251	0.390	356

In probability theory, the probability density function $h(x)$ of the sum of two independent random variables is the convolution of the probability density functions $h_1(x)$ and $h_2(x)$ of the two random variables:

$$h(x) = \int_{-\infty}^{+\infty} h_1(\lambda) * h_2(x-\lambda) d\lambda = h_1(x) * h_2(x) \tag{23}$$

The utility of this function can be attributed to the fact that calculating convolution allows for the simple determination of the probability density function of the sum of independent random variables. This is very useful for understanding and analyzing complex probability distributions, calculating the expected values and variances of the sum of random variables, and so on. By using convolution operations, the paper can combine the probability density functions of two independent random variables to obtain the probability density function of their sum. This approach enables scholars to more readily study and describe the combined distribution of several random variables and extract valuable information from it. From the perspectives of statistics and applications, this is of great significance for simulation, prediction, and decision-making problems. In summary, using the convolutional function of the probability density function of two independent random variables can conveniently calculate the probability density function of the sum of random variables, providing convenience for scholars to study the sum of random variables in probability theory and statistics.

If X_1 and X_2 are evenly distributed in the angle space $[0,2\pi)$ and X is the sum of X_1 and X_2, the probability density function of X is triangular in the range of $[-2\pi,2\pi)$, because the convolution of the two rectangular functions is a triangular function.

3.4 Mobile Robot Vision Image Feature Recognition

Based on the VC dimension theory of statistical learning theory and the principle of structural risk minimization, machine vision seeks the best compromise between the complexity of the model (i.e., the learning accuracy of specific training samples) and the learning ability (i.e., the ability to identify any samples without errors). The aim is to obtain the best generalization ability. Machine vision has been widely utilized by scholars across various fields due to its numerous benefits, including sample prioritization, algorithm simplification into a quadratic problem, algorithm complexity not being dependent on sample dimension, avoiding the " dimension disaster " problem, and simplification of classification and regression problems, as well as good robustness. In this paper, the feature extraction of mobile robot visual image has been discussed, which paves the way for image classification. This chapter mainly realizes the image classification and recognition by programming the machine vision algorithm.

After the program starts, the training sample data of the image is read. The sample space size is a certain value, which is set to 50, 100 or 150 in this paper. When the program determines that all the samples are read, the feature extraction of each image is started.

First, the image color feature extraction mainly uses the image histogram feature and the image color feature after histogram equalization and establishes the sample color feature database.

Second, the sample image is grayed to prepare for image texture feature extraction. Graying adopts $f(i,j)=0.3R(i,j)+0.59G(i,j)+0.11B(i,j)$, in which $f(i,j)$ is the grayscale values of a pixel after conversion. $R(i,j)$、$G(i,j)$、$B(i,j)$ are the sizes of red, green and blue primary colors of the original image respectively.

Then, the texture feature of the sample image is extracted, and the texture feature database is established. Here, the texture feature is extracted by the method of image gray level co-occurrence matrix.

Next, two methods are used to establish the support vector machine feature database. One is to use the overall color histogram and texture feature as the feature vector, and the other is to use the three primary color histogram and texture feature as the feature vector. The next step is featuring training. The sample features of four road images are used as training sets for support vector machine feature training. Through rigorous training, a support vector is derived that is capable of matching the sample data features of each image to the largest extent possible. This support vector can serve as a foundation for machine vision to accurately classify the diverse visual features of mobile robot images. After the establishment of sample space features and feature vectors, read in the image data in the test database, extract color and texture features, and then carry out classification and recognition. When a certain data meets certain classification requirements, it will be classified into this class. When the data cannot be classified into any image class, the features of the image will be returned to the feature learning part. The class of the image is determined by learning. The feature parameters of the image are added to the feature vector of the class to provide data support for the subsequent establishment of machine vision model.

Finally, through the classification of Lowe algorithm, the purpose of mobile robot visual image feature recognition is achieved.

The process of image feature recognition for mobile robot vision based on machine vision is shown in Figure 4.

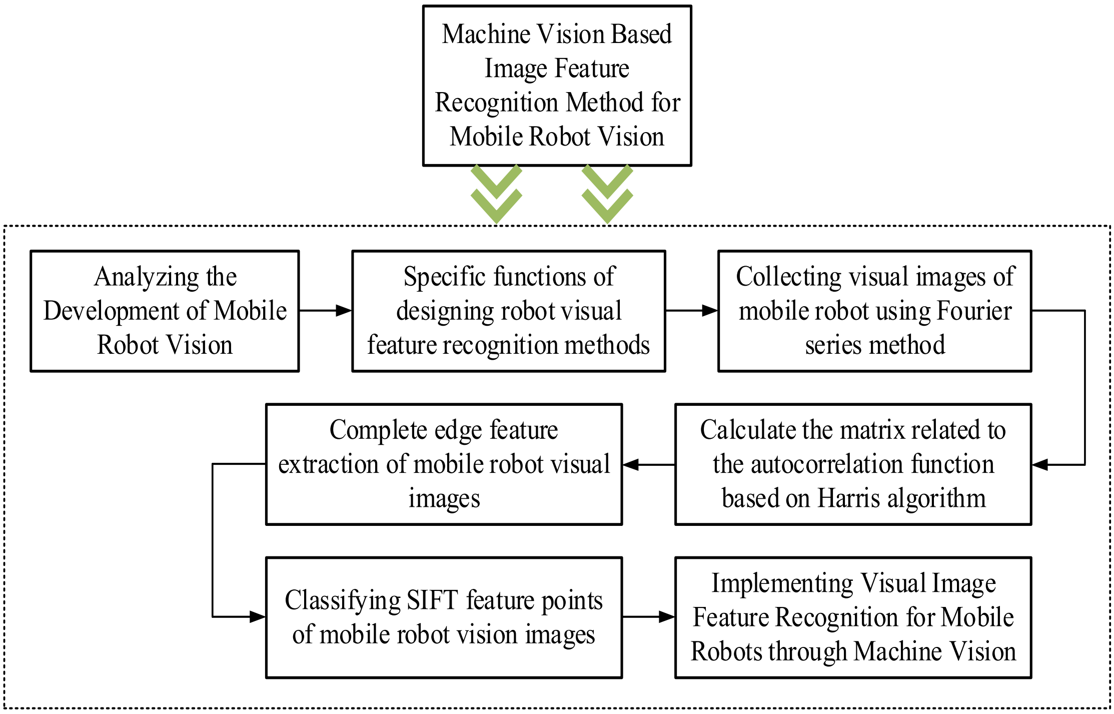

Figure 4: Process of Image Feature Recognition Method for Mobile Robot Vision Based on Machine Vision

4. Experiment

4.1 Experimental Design

The results section first designed the experiment and analyzed and evaluated the image recognition accuracy, recognition time, extracted loss values, and recognition noise to better illustrate the effectiveness of machine vision based mobile robot visual image feature recognition.

The visual image feature recognition of mobile robot is completed on the simulation software of MATLAB, and the accuracy and time of visual image feature recognition of mobile robot are verified. Among them, Python will be used as a programming tool, and the operating system will utilize Windows XP. After the vision system completes the positioning of the target workpiece, the workpiece position information obtained in the camera is transformed into the coordinates of the robot world coordinate system in the form of matrix transformation. Since image feature recognition cannot be carried out during positioning, it can only be performed after the workpiece has traveled a certain distance on the assembly line. When the positioning is completed, the encoder value of the motor on the pipeline can be cleared. The specific experimental robot image sample is shown in Figure 5.

Figure 5: Mobile robot

4.2 Result analysis

4.2.1. Image recognition accuracy

To verify the accuracy of mobile robot visual image feature recognition under different methods, experiments are carried out by using visual attention recognition method [7], multi visual feature acquisition and fusion recognition method [8], visual perception and tactile prior knowledge learning and recognition method [9] to compare with the research method. The results are shown in Table 2.

Table 2. Image recognition accuracy under different methods

Number of Images / pieces	Accuracy of visual image feature recognition for mobile robot			
	Visual attention recognition method	Multi vision feature fusion recognition method	Visual perception learning recognition method	Research method
100	76.56%	72.23%	68.32%	98.43%
200	69.83%	76.36%	69.76%	99.35%
300	70.25%	71.35%	68.61%	98.65%
400	73.65%	72.46%	59.62%	99.52%
500	75.52%	77.82%	72.98%	97.54%
600	78.54%	72.15%	66.22%	96.98%

According to Table 2, when the number of images was 600, the accuracy of mobile robot visual image feature recognition of visual attention recognition method was 78.54%. The accuracy of mobile robot visual image feature recognition of multi visual feature fusion recognition method was 72.15%. The accuracy of mobile robot visual image feature recognition of visual perception learning recognition method was 66.22%. The accuracy of mobile robot visual image feature recognition was 96.98%. Under this method, the accuracy of mobile robot visual image feature recognition was much higher than other methods, which showed that this method had high image feature recognition accuracy.

4.2.2 Robot vision image feature recognition time

To verify the time for mobile robot visual image feature recognition under different methods, experiments were carried out by comparing visual attention recognition method [7], multi visual feature acquisition and fusion recognition method [8], visual perception and tactile prior knowledge learning and recognition method [9] and research method. The results are shown in Figure 6.

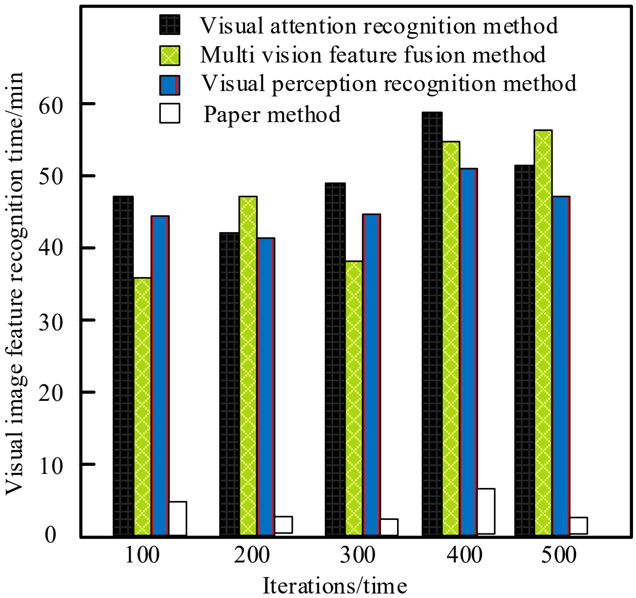

Figure 6: Time for robot visual image feature recognition under different methods

By analyzing Figure 6, there were differences in the time of robot visual image feature recognition under different methods. When the number of iterations was 500, the time for visual image feature recognition of visual attention recognition method was 51 min. The time for visual image feature recognition of multi visual feature fusion method was 56 min. The time for visual image feature recognition of visual perception recognition method was 45 min. The time for visual image feature recognition of research method was 3 min. The robot visual image feature recognition time of the research method was significantly lower than that of the other two methods, which showed the highly robot visual image feature recognition efficiency.

4.2.3 Loss value of edge feature extraction of robot vision image

To verify the loss value of edge feature extraction of mobile robot visual image under different methods, experiments and recognition methods [9] and the research method were carried out using visual attention recognition method [7], multi visual feature acquisition and fusion recognition method [8], visual perception and tactile prior knowledge learning. The results are shown in Table 3.

Table 3. Loss value of image edge feature extraction under different methods

Number of Images / pieces	Loss value of edge feature extraction in mobile robot vision image			
	Visual attention recognition method	Multi vision feature fusion recognition method	Visual perception learning recognition method	Research method
100	2.17%	2.28%	3.24%	1.23%
200	3.67%	4.23%	5.37%	2.14%
300	5.18%	6.39%	7.24%	3.33%
400	6.24%	6.75%	7.38%	4.21%
500	8.07%	8.61%	9.37%	5.17%
600	8.96%	9.16%	9.62%	6.38%

According to Table 3, when the number of images was 600, the loss value of edge feature extraction of visual attention recognition method for mobile robot vision image was 8.96%. The loss value of edge feature extraction of multi visual feature fusion recognition method for mobile robot vision image was 9.16%. The loss value of edge feature extraction of visual perception learning recognition method for mobile robot vision image was 9.62%. The loss value of edge feature extraction of mobile robot vision image was 6.38%. Under the research method, the loss value of edge feature extraction of mobile robot vision image was far lower than that of other methods. This showed that the loss value of edge feature extraction of research method was small.

4.2.4 Noise evaluation of robot vision image feature recognition

To evaluate the noise of mobile robot visual image feature recognition under different methods, experiments and recognition methods [9] and research method were carried out using visual attention recognition method [7], multi visual feature acquisition and fusion recognition method [8], visual perception and tactile prior knowledge learning. The visual image feature recognition of mobile robots was completed on MATLAB simulation software, and the noise situation of visual image feature recognition of mobile robots was verified. A total of 50 evaluations were conducted, and the average error value and error equation average results of noise evaluation were obtained. Among them, Python will be used as a programming tool, and the operating system will utilize Windows XP. The results are shown in Table 4.

Table 4. Noise Evaluation of Image Feature Recognition Under Different Methods

	Noise evaluation of mobile robot vision image feature recognition				
	White noise	Visual attention recognition method	Multi vision feature fusion recognition method	Visual perception learning recognition method	Research method
Mean value of error	1	1.35	1.37	1.39	113
	2	1.52	1.61	1.67	1.45
	3	1.87	1.95	2.08	1.73
	4	2.39	2.45	2.59	2.24
	5	3.29	3.38	3.42	2.68
	6	4.08	4.18	4.75	3.08
Error	1	0.10	0.12	0.13	0.07

variance	2	0.15	0.19	0.25	0.14
	3	0.26	0.28	0.36	0.23
	4	0.34	0.35	0.37	0.32
	5	0.39	0.41	0.46	0.38
	6	0.43	0.45	0.53	0.41

To enhance the precision and steadiness of the technique, Gaussian noise with zero mean and standard deviation ranging from 1 to 6 is incorporated into the image. From Table 4, under different noise conditions, the mobile robot vision image feature recognition noise was the smallest, regardless of the error mean value or error variance, which was more conducive to image feature extraction.

5. Conclusion

This paper presented a feature recognition method of mobile robot visual image based on machine vision. The specific module of mobile robot visual feature recognition was designed. The mobile robot visual image was collected by Fourier series method, and the edge feature extraction of mobile robot visual image was completed according to Harris algorithm. SIFT feature points of mobile robot visual image were classified, and mobile robot visual image feature recognition was realized through machine vision. The following conclusions could be drawn from the experiment:

a. When the number of images was 600, the accuracy of mobile robot visual image feature recognition was 96.98%. It showed that the proposed method had high accuracy of image feature recognition.

b. When the number of iterations was 500, the visual image feature recognition time of this method was 3 min, indicating that the robot visual image feature recognition efficiency of this method was high.

c. When the number of images was 600, the loss value of mobile robot vision image edge feature extraction was 6.38%. It showed that the research method had lower loss value of image edge feature extraction.

d. Under different noise conditions, the mean and variance of the error in mobile robot visual image feature recognition were the lowest, which showed the low image noise of research method.

In summary, improved visual feature recognition technology can make robots more intelligent, improve their autonomy and efficiency, and enable robots to better understand and adapt to human environments, especially in the fields of service robots and collaborative robots. At the same time, this study will promote technological innovation, with significant economic benefits and broad social impacts.

References

[1] J.C. Huang, Q. Shu, X.C. Zhu, L. Zhou, H.Z. Liu, J. Lin, “Robot Visual Recognition and Sorting Strategy Based on Transfer Learning”, Computer engineering and Application, Vol.55, No.08, pp. 232-237, 2019,

[2] S.U. Liu, G.H. Gu, J.Z. Wang, X.J. Cao, “Baxter Target Recognition Algorithm for Cooperative Robot Based on Machine Vision, Electro optic and control”, Vol.26, No.04, pp. 95-99, 2019.

[3] W. Zuo, C.L. Li, L.J. Yu, Z.B. Zhang, R. W. Wang, X.G. Zeng, Y.X. Liu, Y.Y. Xiong, “Shadow–Highlight Feature Matching Automatic Small Crater Recognition Using High-Resolution Digital Orthophoto Map from Chang’ E Missions”, Journal of Geochemistry, Vol.38, No.4, pp. 14-21, 2019.

[4] J.Y. Ma, T.J. Zhang, G.D. Jing, W.J. Yan, B. Yang, “Ground-Based Cloud Image Recognition System Based on Multi-CNN and Feature Screening and Fusion”, IEEE Access, Vol.8, No.21, pp. 2-18, 2019.

[5] F. Özyurt, “Efficient Deep Feature Selection for Remote Sensing Image Recognition with Fused Deep Learning Architectures”, The Journal of Supercomputing, Vol.76, No.4, pp. 1-19, 2020.

[6] L.P. Gao, X.L. Zhang, J.P. Gao, S.X. You, “Fusion Image Based Radar Signal Feature Extraction and Modulation Recognition”, IEEE Access, Vol.16, No.28, pp. 65-79, 2020.

[7] Y. Kong, H. Liang, Q. Zhang, Human Behavior Recognition Based on Visual Attention, Computer System Application, Vol.28, No.05, pp. 42-48, 2019.

[8] J. Zeng, D. Du,Automatic Recognition of Weld Bead Trajectory Based on Multi Vision Feature Acquisition and Fusion, Journal of Mechanical Engineering,Vol.55, No.07, pp. 127-136, 2019.

[9] T. Xue, W.H. Liu, Z.Y. Pan, W.M. Wang, “Robot Stable Grasping Based on Visual Perception and Tactile Prior Knowledge Learning”, Robot, Vol.43, No.01, pp. 1-8, 2021.

[10] L. Hong, “Research on Pattern Recognition Method of Machine Vision”, Modern Information Technology, Vol.4, No.11, pp. 83-85, 2020.

[11] S. Ying, Y. Weng, B. Luo, G. Li, B. Tao, D. Jiang, D. Chen, “Gesture Recognition Algorithm Based on Multi-scale Feature Fusion in RGB-D Images”, IET Image Processing, Vol.32, No.21, pp. 123-138, 2020.

[12] W. Wang, C. Tang, X. Wang, Y.H. Luo, Y.L. Hu, J. Li, “Image Object Recognition via Deep Feature-Based Adaptive Joint Sparse Representation”, Computational Intelligence and Neuroscience, Vol.29, No.2, pp. 1-9, 2019.

[13] C.W. Peng, Z. Yang, “Image Feature Extraction and Object Recognition Based on Vision Neural Mechanism”, International Journal of Pattern Recognition and Artificial Intelligence, Vol.34, No.06, pp. 3340-3342, 2020.

[14] J. Yao., C. L. Williams., Hussain. F, Kouri. D J. “Generalized Fourier Transform Method for Solving Nonlinear Anomalous Diffusion Equations”, Vol.10, No.12, pp.1039-1047, 2019.
[15] Y. Zhu, Y. Jiang, “Optimization of Face Recognition Algorithm Based on Deep Learning Multi Feature Fusion Driven by Big Data – ScienceDirect”, Image and Vision Computing, Vol.104, No.21, pp. 65-78, 2020.
[16] Y. Wu, J. Zhao, N. Yu, “Indoor Surveillance Video Based Feature Recognition for Pedestrian Dead Reckoning”, Expert Systems with Applications, Vol.21, No.21, pp. 55-68, 2021.
[17] Y. Zhu, W. Min, S. Jiang, “Attribute-Guided Feature Learning for Few-Shot Image Recognition”, IEEE Transactions on Multimedia, Vol.16, No.29, pp. 6-18, 2020.
[18] Shi. Y, “Image recognition of competitive aerobics movements based on embedded system and digital image processing”, Microprocessors and Microsystems, Vol.82, No.Apr., pp.103925.1-103925.6.
[19] J. Dong, C. Zhang, Y. Wang, “Visual localization Experiment for Robot Intelligent Grasping Task”, Experimental technology and management, Vol.37, No.03, pp. 56-59, 2020.
[20] Y. Wang, Y.R. Li, “Simulation Research on Dynamic Collision Detection Method between Multi Robot Bodies”, Computer Simulation, Vol.37, No.04, pp. 335-339, 2020.
[21] Zhangfang Hu, Lan Wang, Yuan Luo, Yanling Xia, and Hang Xiao, "Speech Emotion Recognition Model Based on Attention CNN Bi-GRU Fusing Visual Information," Engineering Letters, vol. 30, no.2, pp427-434, 2022.
[22] Yi-Cheng Lee, Yu-Chen Wu, and Syh-Shiuh Yeh, "Development of an On-Machine External Thread Measurement System for CNC Lathes Using Eye-in-Hand Machine Vision with Morphology Technology," Engineering Letters, vol. 29, no.3, pp901-912, 2021.