Computer Vision-based Intelligent Bookshelf System

Hao Yu1,*

1 School of software, Zhengzhou University, Zhengzhou, China

*Corresponding author e-mail: 2604962074@qq.com

Abstract. This paper is devoted to the research of an intelligent bookshelf system. By introducing the current status of reading and purchasing books by people in various countries around the world, the value of the existence of the intelligent bookshelf is analyzed and proved. By introducing the main related technologies and algorithms used in the intelligent bookshelf system, this paper brings out the introduction to the principle and mode of operation, and an analysis of the system.

1. Research background

With the improvement of people's cultural level, more and more people realize the importance of reading. According to statistics, the top five countries in the world in terms of reading per capita are: Russia 54 books, Israel 50 books, Germany 47 books, Japan 45 books, and Austria 43 books. The world's per capita book purchase data are: Israel 64 books, Russia 55 books, and the United States 50 books. Borrowing books, buying books and reading books has become a habit of people's life, but when borrowing books or buy books, the trouble of finding books is really a headache for book lovers and book sellers.

This paper is dedicated to researching an intelligent bookshelf system. The core of the system uses computer vision technology and deep learning technology, combined with speech recognition technology, speech synthesis technology, and embedded technology. The intelligent bookshelf system makes the process of buying books or borrowing books from the library easier and faster, and also makes people who have a lot of books in their homes find books more quickly and efficiently when they want to read. This system can help readers who need to find a specific book to quickly locate the precise position of the book on the bookshelf, and at the same time generate statistical data of each book through the reader's purchase and reading habits, and provide a basis for bookstores and libraries to purchase and sell books. Utilizing this intelligent bookshelf system, the bookstore or library does not even need to categorize books. It only needs to put the books on the bookshelf. The intelligent bookshelf system can intelligently classify and store books.

Nowadays, countries are paying more and more attention to artificial intelligence technology, and various artificial intelligence technologies are becoming more and more mature. More and more artificial intelligence technologies are moving towards commercialization and industrialization. More and more traditional things are embedded in artificial intelligence. It is the general trend to provide a more convenient and efficient experience. Bookshelf is no exception. Bookshelf also needs a better intelligent system to give it its soul.

2. Analysis of research status at home and abroad

The problems facing smart bookshelves are similar to the "Who-Take-What" problem in unmanned retail scenarios. At present, smart bookshelves in domestic and foreign markets are also designed
around this problem. The proposed smart bookshelf should not only solve the "Who-Take-What" problem, but also the "How-Easy-Take" problem.

At present, the intelligent bookshelf on the market can be regarded as an unmanned retail bookshelf, but its essence still belongs to the unmanned retail scene, which is mostly used for the sale of a small number of books. The smart bookshelf proposed in this article is used in bookstores and libraries with huge collections.

At present, the existing intelligent bookshelves on the market use computer vision technology to solve the "Who" problem, such as face recognition. This technology is indeed a good solution for a single retail bookshelf, but what this article describes the intelligent bookshelf system is mainly used in bookstores and libraries. There are a large number of books in the smart bookshelf system. The face recognition scheme for borrowing books and purchasing books will inevitably increase costs. Therefore, this system uses a corresponding APP to bind Alipay, WeChat or other mobile payment software for identification. The "What" problem is the use of computer vision technology and binocular vision for book recognition. For the "How-Easy-Take" problem, the system uses speech recognition, speech synthesis, binocular vision, and mobile APP to solve the problem.

3. Related technologies

3.1. Binocular vision

Binocular Stereo Vision is a very important technology in the field of computer vision. At present, the main application fields are trajectory tracking, robot obstacle avoidance, autonomous cruise and flow statistics. Binocular vision uses two or more cameras to calculate the depth of an image based on their relative positions. Depth refers to the distance from a point on the image to the camera lens. To briefly introduce the binocular vision algorithm, in this article, we only consider the case of two cameras. Assume that the camera on the left is VL and the camera VR on the right. In order to measure the depth information of the scene, we need to have several First, the relative position of the two cameras is a PNP problem. Now there is mature camera calibration software that can easily obtain the relative positions of the two cameras and estimate the relative error. Second, the reality the pixel position of a point in the world is known, which is to obtain a pair of matching points. The matching points in the pair of pixels correspond to an actual position point in the real world. Generally, this type of algorithm is reduced to a stereo matching algorithm. For each pixel $p_l = (x_l, y_l)$, search for a matching $p_r = (x_r, y_r)$ in VR. Matching means that PL and PR come from the same point P in the real world. This point is called a scene point.

Usually in computer vision, this point is acquired using the nature of epipolar constraints. The matching pixels of each image on another image can only be found on a line called an epipolar line. This property makes the Matching, instead of searching in the entire image, you only need to match on a pixel line, and filter out some points that are easy to mismatch. At this time, if only the “forward parallel” configuration is considered, it can be simplified to a certain degree, but it is difficult to achieve the “forward parallel” configuration in the real scene. Therefore, the principle of stereo correction is generally used to make the two cameras “forward parallel”. Configuration, and finally calculate the image depth using the following formula.

$$d = \frac{fb}{m}$$

Among them, $d$ is the depth of the scene pixel point P, $f$ is the focal length of the camera, $b$ is the baseline of the binocular camera, and $m$ is the distance of the scene point in the coordinates of the two camera pixels [1].

3.2. Image Identification

Image recognition is a technology that uses a computer to process, analyze, and understand images to identify targets and objects in various modes. It is a practical application of deep learning algorithms. At this stage, image recognition technology is generally divided into face recognition and product recognition. Face recognition is mainly used in security inspection, identity verification and
mobile payment; product recognition is mainly used in the circulation of goods, especially unmanned shelves, intelligent retail cabinets, etc. Unmanned retail.

The image recognition algorithm used in the intelligent bookshelf described in this article is the SIFT algorithm. This algorithm was proposed by Lowe in 2004. This algorithm uses the convolution of the original image and the Gaussian kernel to establish the scale space and extracts it on the Gaussian difference space pyramid. Feature points for scale invariance. This algorithm has certain affine invariance, perspective invariance, rotation invariance and illumination invariance, so it has been most widely used in improving image features.

The algorithm can be roughly divided into three steps: 1) construction of Gaussian difference pyramid; 2) search of feature points; 3) feature description.

In the first step, it uses the structure of groups and layers to build a linear pyramid structure, which allows us to find feature points on continuous Gaussian kernel scales. It uses first-order Gaussian difference to approximate the Gaussian Laplacian kernel, which greatly reduces the amount of calculation.

In the second step of the feature point search, the main key step is the interpolation of extreme points, because in a discrete space, the local extreme points may not be the extreme points in the true sense, and the true extreme plant points can fall in the gap between discrete points. So we need to interpolate these gap positions, and then find the coordinate positions of the extreme points.

The other key link in the second step is to delete the points of the edge effect, because it is not enough to ignore those points where the DoG response is not enough. The value of the DoG will be affected by the edges. Although the points on the edges are not spots, they are not DoG response is also strong. So we have to delete this part. We use the feature that spans the edge and exhibits maximum and minimum principal curvature in the direction of the edge and the direction of the vertical edge. So by calculating the ratio of the principal curvature at the feature point, it can be distinguished whether it is on the edge. This point can be understood by referring to the method of Harris corner.

The last step is the feature description of the feature points. The method of finding the direction of the feature point is to perform histogram statistics on the gradient directions of points in the neighborhood of the feature point. The direction with the largest proportion in the histogram is selected as the main direction of the feature point, and a secondary direction can be selected. When calculating the feature vector, the local image needs to be rotated in the main direction, and then entered into the gradient histogram statistics (4x4x8) in the neighborhood.

3.3. Embedded technology

The chip used in the hardware part of the intelligent bookshelf system studied in this article is a high-performance DSP chip FT-M7002, which uses multi-core, SIMD and VLIW technologies, and optimizes various aspects, including: parallel memory access optimization during storage, Instruction parallel optimization, data level parallel three aspects. The memory access during storage optimizes the DSP-based multi-level memory structure; the instruction parallel optimization of the VLIW technology based on the DSP architecture enables more instructions to be executed at the same time, and plays the role of multiple independent functional units. Soft flow, reduce the idle cycle of the CPU, and achieve concurrent read and execution of multiple instructions; data-level parallel DSP-based SIMD technology, using SIMD technology to give full play to the data parallel processing capabilities of vector units, complete the parallel processing of data, and combine multi-core Parallel execution of multi-core tasks [2].

4. System introduction

4.1. System structure

The intelligent bookshelf system studied in this article is mainly composed of a hardware terminal, a server, and a corresponding applet. As shown in Figure 1, the terminal is an intelligent bookshelf. The intelligent bookshelf is mainly responsible for storing books, identifying books, and realizing the precise positioning of the books; Responsible for the virtual classification of the books identified by
the intelligent bookshelf, and the precise location of the books is stored in the database server; applets are convenient for users to search for books and find the location of the books.

![System structure diagram](image1.png)

**Figure 1.** System structure diagram.

The hardware design of the intelligent bookshelf (Figure 2) is similar to the ordinary bookshelf. It has a rectangular parallelepiped shape and uses a layered structure. Each layer will have two tracks, one on the front and one on the back. These two cameras can pass through the track. Slide left and right on this floor to take pictures of the books placed on the floor; another camera is installed at the top of the bookshelf, and in cooperation with the cameras on each floor, binocular vision technology is used to detect whether the user has taken or placed books in front of the bookshelf behaviour, and detect the specific location where the book is placed. After the user operates, the camera of this layer will resample the book of this layer, re-determine its relative position, and finally send the position information to the server.

![Intelligent Bookshelf Structure](image2.png)

**Figure 2.** Intelligent Bookshelf Structure

4.2. Operation principle and process

The main function of the hardware part of the intelligent bookshelf system studied in this article is: without adding any changes to the original books, without adding any electronic tags to the books, and accurately positioning the position of each book on the bookshelf. In order to achieve this function, this system uses binocular vision and image recognition technology. The technical principle is described in the relevant technical section; the specific operation process of the hardware section is:

1) According to the binocular vision-related algorithm, determine whether someone has taken a book from the bookshelf or placed a book on the bookshelf in front of the smart bookshelf. If not, the system continues to be in the detection state, and if so, proceed to the next step.

2) According to the relevant algorithm of binocular vision technology, the position where the user picks up or puts the book is detected. The two cameras in the layer scan the left and right twice, and sample the position information of the book to determine each book in the layer. Book name and relative position information; the sampling method is based on image recognition technology. The book name information can be obtained by identifying the side of the book, and the image's relative position information can be obtained using image segmentation and image recognition.

3) The intelligent bookshelf sends the title information and the relative position information of the books (the bookshelves where the books are located, the number of layers on which they are located, and the relative positions on this layer) to the server.
The smart bookshelf system studied in this article has special customization for home scenes. In home occasions, the hardware part of the smart bookshelf system adds speech recognition and speech synthesis functions. Users can directly input the books they want to find when they pick up the books. The intelligent bookshelf can inform users of the specific location of the books by voice. In public places such as bookstores and libraries, the smart bookshelf system uses applets to work with the smart bookshelf. Users need to enter the books they need to find on the applet terminal, and the applets will show users the specific location information of the books.

The intelligent bookshelf system studied in this article uploads various information about books on the bookshelf to the server, and the server can classify the books virtually. Therefore, the intelligent bookshelf system studied in this article does not need to classify and store books.

5. System analysis
The intelligent bookshelf system studied in this paper combines binocular vision technology, image recognition technology and embedded related technologies. The binocular vision technology is mainly used to detect whether someone is holding or putting a book. The technology is currently mature. The recognition accuracy of the technology has reached 98.6%, and the system delay is 3.82 milliseconds.

The system studied in this paper uses image recognition technology for the positioning of books. This system relies on the side of the book to identify the title and the relative position of the book. According to our sample survey of some libraries and bookstores, there are some books, and their sides cannot be used to identify the book title information. At this time, an error will occur in the system. In order to solve this unavoidable error, The bookshelf system adds the function of reporting errors to the applet when an error occurs. The system will indicate the specific location of the book that cannot be identified. At this time, the user needs to manually enter the name of the book to record in the system. Such a situation will not cause very serious Great impact. According to sample statistics, the sides are blank, and books that cannot identify the title information account for only 0.02% of all books.

6. Conclusion
The hardware side of the intelligent bookshelf system studied in this paper uses the FT-M7002 chip. Based on this chip, various aspects have been optimized to enable efficient implementation of image recognition and binocular vision related algorithms, and achieve accurate positioning of books on the bookshelf. Based on the image There are some flaws in the identification and positioning method. When there is no title information on the side of the book, the book cannot be identified and located, but after sampling statistics, it is known that the probability of this phenomenon is not high, and it can be solved by adding relevant algorithms. The advantages of the book positioning method based on image recognition are obvious. All books can be positioned without adding any electronic tags, which can effectively reduce costs and manpower.

7. Acknowledgments
This work was financially supported by Zhengzhou University Student Innovation and Entrepreneurship Training Program fund.

References
[1] Guan Jieqiong, Research and Implementation of Object Scout Recognition Algorithm Based on Computer Vision, Shanghai Normal University Master of Philosophy, 2019.11.
[2] Wang Rui, Embedded Image Recognition System and Optimization Based on Convolutional Neural Network, Master Thesis,2019.6.
[3] Zhang Yuting, Zhu Minling, Zheng Zhiwei, MonocularBased on the Analysis and Realization of Binocular Vision Mapping Process, Academic trends and new perspectives, 2019.
[4] Mao Mengna, Research on Target Recognition and Grab Technology of Binocular Vision Robot, Master Thesis, 2019.6.4.
[5] Lian Zifeng, Research on Image Recognition Algorithms Based on Deep Neural Networks, PhD thesis,2017.3.12..