Visual module based on VSLAM

Letian Xiong¹, Guojiao Liao², Shiqi Xiong³

¹School of software, Jiangxi Normal University, Jiangxi 330003, China
²Civil engineering, Chengdu University of technology, Sichuan 610000, China
³School of software, Jiangxi Normal University, Jiangxi 330003, China

Corresponding author and e-mail: Letian Xiong, 2235266059@qq.com

Abstract. With the rapid development of science and technology, VSLAM (real-time positioning and map construction technology) has been widely used in AR/VR, robots, drones and unmanned aerial vehicles and other fields because of its unique theory and application value in the era of artificial intelligence. However, there are still difficulties on indoor and outdoor positioning navigation and image scene understanding, such as auxiliary rescue after disaster, inspection of large factories, etc. In addition, the combination entity of VSLAM and artificial intelligence has not realized the popularization of consumption level. A vision system positioning module based on VSLAM technology can be transferred to be the carrier of any intelligent robot, which can be used in post-disaster auxiliary rescue, intelligent inspection of large factories, intelligent logistics sorting field and the construction of intelligent city in the future. People can save rescue time, help save life, reduce economic loss and the safety risk of rescue personnel on the auxiliary disaster rescue and intelligent inspection with it. At the same time, it can help achieve intelligent logistics sorting (and its scope expansion) by setting up network with 5G (5th generation wireless systems) in logistics community distribution.

1. Introduction

This vision system module is completed by using the semantic slam technology independently developed. It is directly aimed at Poor GPS signal environment indoor, intelligent robot can to solve the problem of inaccurate positioning, assisting human to locate and navigate. VSLAM technology is introduced to realize 3D scene reconstruction and timely map construction in the indoor environment with poor GPS signal. It has the outstanding function of real-time and fast modeling, precise positioning and independent control, which effectively solves the problem of failure or inaccuracy of conventional positioning methods of intelligent machine in the environment with weak GPS signal (such as indoor environment). Through the technology accumulation of depth vision technology, combined with the mature in-depth learning algorithm and industry data based on CNN (convolutional neural network), the vertical application image understanding algorithm is developed. The innovation of this algorithm is to use the advantages of depth image different from traditional image, and obtain more accurate image language understanding through the acquisition of higher-dimensional feature point set And cognition. Its platform can effectively serve the modeling potential market in the fields of disaster assistance, intelligent monitoring and security, intelligent logistics sorting and future smart city construction.

2. Text organization
2.1. Development of machine vision

Machine vision technology is one of the core technologies to realize computer integrated manufacturing, which can be used to realize precision control, intelligence and automation of equipment. Compared with human vision, machine vision has unique advantages, it can simultaneously receive information and extend it to the processing and judgment of information. With the improvement of supporting infrastructure, the expansion of the overall scale of manufacturing industry, and the improvement of intelligent level and other factors, the demand of Chinese machine vision market shows an overall upward trend [1]. Through a series of data, we can see that machine vision has great market development potential in Chinese market, which actually requires a large number of innovative forces to pour in.

2.2. Market analysis of machine vision

Recently people are getting more and more familiar with 5 G communication environment, 5 G is creating an efficient and convenient communication network, which provides convenience to help machine vision system to connect unconnected data to the central system of the factory to realize real-time interaction between data by using 5G, at the same time, it promotes the ability of cloud load processor, enhance algorithm processing ability, realizing more efficient processing of detection task effect, providing more possibilities and unknowns for machine vision applications [2], and enhance the plasticity of machine vision market.

China's machine vision market is mainly divided into three parts: international integrated automation industry company, international specialized machine vision manufacturer and domestic specialized machine vision company. Among them, foreign-funded enterprises occupy the dominant position in the market, while China's relatively small number of developers are more represented by product agents, system integration and equipment manufacturing, while the number of the grass-roots level developers is less [3].

| Corporate Name | Country       | Market Share |
|----------------|---------------|--------------|
| Cognex         | U.S.A         | 39%          |
| KEYENCE        | Japan         | 25%          |
| Baumer         | Switzerland   | 3%           |
| Dalsa          | Canada        | 2%           |
| NT             | U.S.A         | 2%           |
| other          | Russia, China, France | 29% |

Our team developed a precise positioning and path-finding navigation system based on VSLAM technology and intelligent machines. Currently, the competitors in the machine vision industry are developing unevenly. At present, the well-known enterprises domestic and overseas are Daheng image, Cognex, KEYENCE, Baumer and other limited companies, which basically include most of products on the market. The team carried out a corresponding detailed investigation, and compared the product and reached a conclusion as shown in the following table:

| Performance indicators | Daheng image | Cognex | KEYENCE | Baumer | Machine vision module |
|------------------------|--------------|--------|---------|--------|-----------------------|
| Response time (μs)     | 15           | 20     | 13      | 17     | 10                    |
| Obstacle range         | 0.7-15       | 2-25   | 0.5-20  | 0.7-20 | 0.5-15                |
| Standby time(h)        | 28           | 40     | 50      | 25     | 35                    |
Horizontal positioning accuracy (m)  
2.5  1  2  0.8  0.01

Vertical positioning accuracy (m)  
0.5  0.5  0.5  0.6  0.01

|                  | Hawkvis | Kuangshi Technology | Visual module |
|------------------|---------|---------------------|---------------|
| Autonomous obstacle avoidance accuracy | 2       | 1                   | 0.05          |
| Horizontal positioning accuracy (m)     | 0.5     | 0.8                 | 0.01          |
| Vertical positioning accuracy (m)       | 0.5     | 0.6                 | 0.01          |
| Training model                         | 3540    | 6877                | 25014         |
| The data transfer                       | 3Mbps   | 2 Mbps              | 1.5 Mbps      |
| The processing time                     | 1s      | 1.5s                | <1s           |
| Misjudgment rate                        | 7%      | 5%                  | 1.37%         |
| False negative rate                     | 1%      | 2%                  | 1%            |

### Table 3. Comparison of scene understanding algorithms of each enterprise image.

3. VSLAM based visual module design

3.1. Functional design

After 3D scene reconstructed by using VSLAM technology and enter the target location, we can find the best route according to the data which has been collected by using A* algorithm. The point cloud data is compressed and uploaded to the cloud server, where depth information is received and transformed into point cloud data according to geometric relations. Noise data is filtered through filtering algorithm to effectively retain the clustering characteristics of point cloud model. Then the recognition and segmentation of the ground plane are realized based on the patented algorithm, then project the obstacle onto the ground to calculate obstacle projection point of convex polygons, then int can generate a two-dimensional map in front of your eyes through convex polygons. Finally, the improved A* algorithm is used to calculate the path through. When the video data is only transmitted back by the intelligent robot's structured light sensor, there is a blind area of the machine's field of vision, which may lead to unnecessary collision of the intelligent robot.
By calculating the depth value of each frame of data acquired by using the structured light sensor, when the video data is only transmitted back by the intelligent robot's structured light sensor, there is a blind area of the machine's field of vision, which may lead to unnecessary collision of the intelligent robot. Then it could help average out the data, which is used to judge whether there are obstacles in the uav, and use that as a basis for deciding whether to continue flying. The cloud processing center uses the dynamic ant colony algorithm [4] to obtain the optimal path and assign it to the intelligent robot, when dynamic programming algorithm get the optimal path in the shortest time, optimal results can be obtained with minimal resource consumption. The spatio-temporal data indexing algorithm developed based on r-tree index extension can locate the location of intelligent robots that meet the spatio-temporal range in real time from hundreds of millions of mobile intelligent robots.

3.2. Image scene understanding based on deep learning

Through the depth of visual technology accumulation, combined with the current academic circles have mature based on the depth of the CNN (convolution neural network) learning algorithm, and industry data [5], to develop the application of vertical image understanding algorithms, the algorithm of innovation is to take advantage of the depth image different from traditional image, is obtained by more high-dimensional feature point set to obtain a more accurate image language understanding and cognition. The urban smart eye system USES the deep neural network tensorflow to train a large number of wall damage photos and obtain the evaluation model. After receiving an actual wall damage photo, the model returns 1-10 evaluation results in real time. The larger the returned evaluation results, the more serious the wall damage will be. The specific implementation steps are as follows:

Step 1: in order to train the neural network, we need to obtain a large number of photos of wall damage. The consultant's advice is to manually annotate each graph and combine all the data into one data set.

Step 2: in order to obtain the damaged features of wall pictures, CNN convolutional neural network of tensorflow was used. Convolutional neural network can avoid the problem that the deep neural network model is too large, during the process it can save the matrix characteristics of the picture in the training. We can first extract the characteristics of the wall damage by using the convolutional deep neural network to evaluate the wall damage, including the size, density, distance and width of
cracks. Moreover, the image transmitted back belongs to high resolution, which is specifically reflected in its high quality and large number of pixels. Therefore, we used convolutional neural network to extract the features of the image and give the probability distribution.

Step 3: in order to reduce the burden of the neural network, Opencv technology is used to preprocess the image: ① the image will be uniformly cropped to the size of 64x64 pixels, and the central region will be clipped for evaluation or random clipping for training; ② The image will be approximately whitened, so that the model is not sensitive to the dynamic range changes of the image.

**Figure 2.** Loss function image and learning rate image of neural network.

Step 4: in order to get the model with high accuracy, we are supposed to train the model. As for training, we have adopted a series of random transform methods to artificially increase the size of the data set: ① randomly flip the image left and right; ② randomly varying the brightness of the image; ③ randomly transform the contrast of the image.

**Figure 3.** Training analysis diagram of the system (unit: 100 times).

3.3. Functional innovation and advantages
The combination of the existing fixed-point photographing technology and mapping software is not accurate enough for the location of the disaster scene, with the incomplete coverage of WiFi network, will increase the difficulty of rescue. Under the condition of its own location uncertainty, the system creates a map in the completely unknown environment, using the map for self positioning and navigation at the same time.

People choose those intelligent robot to enter those dangerous areas to assist rescue, when the vision system module receives 3D data and image information collected by intelligent robot structure optical sensor on client side, the robot follows its own planned route and uses improved A* algorithm to calculate traversable path. It is convenient for us to better monitor intelligent robot path finding, ensuring robot work efficiency.
When the visual system module get into the process of scanning modeling, it can use struvisual system module in the process of scanning modeling, take wall photos by using structural light sensor, shooting according to the intelligent robot and cuurent position information to transmit back high-definition photo marks, which can be seen in the three-dimensional model of the wall damage degree mark. After receiving the wall photos, the background server uses the wall evaluation system to make a professional judgment on the damage degree of the wall in the obtained photos, and returns the damage degree of 1-10 grades. Correspondingly, in the 3D point cloud model, the damage degree is marked with numbers, which can directly reflect which place belongs to the highly dangerous area, which is used to prompt the rescue personnel to avoid.

Maskfusion also has outstanding advantages. Maskfusion does not require a static scene, and compared with other dynamic slams, it can enrich the dynamic map with real-time semantic information [6].

4. Conclusions
This vision system positioning module based on VSLAM technology can be transferred to be the carrier of any intelligent robots, which can be used in post-disaster auxiliary rescue, intelligent inspection of large factories, intelligent logistics sorting field and the construction of intelligent city in the future. People can save rescue time, help save life, reduce economic loss and the safety risk of rescue personnel on the auxiliary disaster rescue and intelligent inspection by using it, At the same time it can help achieve intelligent logistics sorting (and its scope expansion) by setting up network with 5G (5th generation wireless systems) in logistics community distribution. With the development of the Deep learning and Machine vision, it is firmly believed that it can provide a more consumer oriented development of new ideas in the construction of future "smart city", help better build a scientific and technological life, and achieve the milestone development of "urban information folding" under the background of the development of science and technology.

Acknowledgement
Time flies, it took almost one year to finish this peper, I learned a lot in the process, thanks to my mentors, they helped me learn many professional knowledge and skills on writing papers, they support me and helped me get more confidence.

References
[1] Machine vision development present situation and development prospect of http://m.elecfans.com/article/988712.html.
[2] 5 g commercial will bring revolutionary change. http://smartcity.asmag.com.cn/gydz/5720.html to machine vision.
[3] Machine vision market research at https://www.jianshu.com/p/f8fb6718db50.
[4] Based on improved ant colony algorithm for robot path planning of http://cdmd.cnki.com.cn/Article/CDMD-10255-1015032614.htm.
[5] Convolution neural network based on the depth of indoor service robot scene understanding technology research at https://xueshu.baidu.com/usercenter/paper/show? Paperid = 1 = xueshu_se t6202r0j00f06709j7x02f09y011833 & site.
[6] MaskFusion: real-time Recognition, Tracking and Reconstruction of Multiple Moving Objects. https://blog.csdn.net/phy12321/article/details/103220226.