Stereo Matching and Image Inpainting Based on Binocular Camera

Yibo Du[1,2,3], Kebin Jia[1,2,3*], Chang Liu[1,2,3]
[1]Department of Informatics, Beijing University of Technology, Beijing 100124 China
[2]Advanced Information Network Beijing Laboratory, Beijing, 100124
[3]Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124 China
[*]E-mail: kebinj@bjut.edu.cn

***Abstract*—Stereo matching is one of the key technologies in the field of computer vision. The depth map obtained by stereo matching contains the three-dimensional information of the scene. The use of depth map is of great significance in the three-dimensional reconstruction of the map and the autonomous navigation of the robot. Aiming at the accuracy and speed of stereo matching, this paper applies a semi-global stereo matching method to match corrected left and right perspective images. Because there are noise points and holes in the matched disparity map, which affect the image quality, a sample block filling method which combines mean filtering and point-by-point scanning is proposed to repair the image. Then a gradient priority selection mechanism is proposed to maintain the edge structure of the object in the process of restoration. Experimental results show that the proposed method is good for the restoration of holes and noises in disparity maps, and the processing speed is improved by about 30% compared with the traditional Criminisi algorithm.**

I. introduction

With the rapid development of robotics industry, computer vision technology has been widely used as an interdisciplinary subject. Binocular stereo matching technology in computer vision is one of the hotspots of research. In recent years, the technology of binocular stereo matching has developed rapidly. Different kinds of depth maps can be obtained by processing images captured by binocular cameras. The depth information contained in each pixel of the depth map reflects the three-dimensional information of the scene, which is of great value to the mapping construction (three-dimensional reconstruction) and ranging navigation of the robot. The sensors for obtaining the depth map mainly include laser radar, binocular camera [1], Kinect depth camera produced by Microsoft and infrared camera produced by FLIR. In this paper, binocular camera is used as the sensor for acquiring images. The advantage of is that the binocular camera is cheap and portable, and the left and right view images are captured from different angles.

Stereo matching method based on binocular camera is a research hotspot in the field of image processing. In 1976, Marr [2] first proposed the application of visual computing theory to binocular stereo matching, which opened up the research of stereo vision. Stereo matching methods are mainly divided into local matching and global matching. Block matching algorithm applied by Wang et al. [3] is a method of local region matching, which adopts fixed SAD window to calculate the matching degree of image blocks between two images. The real-time performance of this algorithm is good, but the quality of the image obtained is not high. Roy et al. [4] first proposed the image segmentation method (GC) which can be applied to stereo matching. It is a global matching algorithm based on graph optimization theory. Based on the information obtained by image segmentation, Vieira et al. [5] proposes a new adaptive optimization method based on Graph Segmentation. The image is segmented and the weight of each region is allocated by adaptive window, which achieves a better refinement effect. S Martull et al. [6] used graph cut to achieve the global matching effect of images. By constructing a global energy function from the established network graph and using the image segmentation method to solve the energy function minimization, a dense disparity image was generated. The image generated by this method is of high quality but time-consuming, which was not conducive to real-time acquisition. In this paper, a semi-global stereo matching method is adopted, which has higher image quality than the local block matching method and lower computational time complexity than the global matching algorithm such as graph cut method, but the quality of parallax image is not perfect. Therefore, in view of the image restoration, the predecessors have also studied from many aspects. In reference [7], EPI light field was applied to depth image restoration, in which the depth image was acquired by RGB-D camera. Fan et al. [8] improved the traditional FMM algorithm to improve the edge preservation of the image, but there will still be a blurred situation. And because the selection of mask image was difficult, the filling effect of the two methods for smaller holes was not obvious. Criminisi et al. [9] proposed an image restoration algorithm based on sample blocks, which improved the restoration effect obviously. However, with the decreasing of confidence value, it is easy to produce false matching of sample blocks, and the repair effect for large-scale damaged areas is poor. On the basis of literature [10], an improved Criminisi image restoration algorithm is proposed. On the basis of the original algorithm, the calculation method of confidence is changed, and the restoration error is reduced by increasing the priority size of the blocks to be filled. However, for the poor quality, the effect of image restoration with a large number of holes is not ideal. Literature [11] started with the point to be repaired, found all the pixels with a certain distance n, and filled them by comparing the similarity between the sample block around each pixel and the

patch to be repaired, thus reducing the repair time. However, the selection of n is difficult to determine, and the restoration of the object structure is not ideal. Liu [12] et al. designed two-layer filters for depth images with holes. The combination of median filtering and pixel filtering can effectively remove noise and isolated holes, but it has poor filling effect for areas with large damaged areas, and the smoothing effect of filtering is more serious.

To solve this problem, a sample block filling method combining filtering operation and pixel-by-pixel scanning is proposed, which can effectively remove independent noise and fill a large range of holes, thus improving the image quality.

The remainder of this paper is organized as follows: Section II mainly applies a semi-global stereo matching method to acquire disparity map by stereo matching of the captured image. Section III proposes a sample block filling method combining filtering operation and pixel-by-pixel scanning to repair the damaged image after matching. Section IV give the experimental results of this paper. By comparing this method with other literature methods, the effectiveness of this method is proved.

II. stereo matching

In this paper, a semi-global stereo matching method is used to match two different disparity images. Firstly, each pixel in the image is matched by block matching method to calculate the disparity between images and draw the disparity map. A global energy function related to disparity map is set up:

$$E(D)=\sum_{p}(C(p,D_p)+\sum_{q\in N_p}P_1 I[|D_p-D_q|=1] + \sum_{q\in N_p}P_2 I[|D_p-D_q|]>1) \quad (1)$$

Where p, q represents a pixel in an image, Np refers to the adjacent pixel point of the pixel p, and C (p, Dp) refers to the cost value of the pixel when the disparity of the current pixel point is Dp. If the parameter in the I[.] function is true, it will return 1, otherwise it will return 0.

Finally, The disparity optimization effect of each pixel is achieved by minimizing the energy function. Each pixel of an image has eight adjacent pixels. Considering the direction from left to right, the disparity value of each pixel is only related to the adjacent pixels on the left. Its function is as follows:

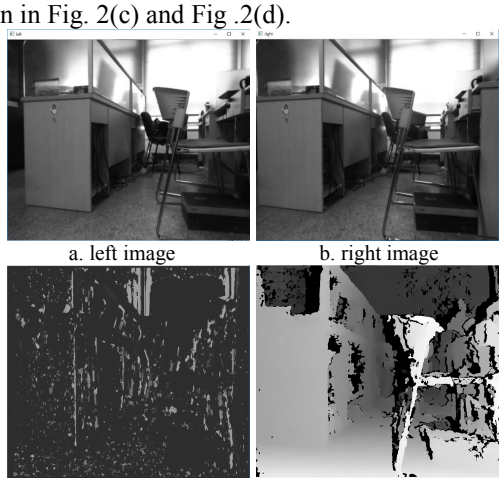
$$L_r(p,d)=C(p,d)+\min[L_r(p-r,d),L_r(p-r,d-1)+P_1,L_r(p-r,d+1)+P_1,\min_i L_r(p-r,i)+P_2]-\min_k L_r(p-r,k) \quad (2)$$

Among them, **r** represents the current direction of the pixel p. And P1 is a penalty coefficient, which applies to those pixels whose disparity value is 1 difference between the pixel P and adjacent pixels. P2 applies to those pixels whose disparity value is greater than 1. $L_r(p,d)$ represents the minimum cost when the disparity value of the pixel P is d from left to right. The formula for calculating $C(p,d)$ is as follows:

$$C(p,d)=\min((d(p,p-d,I_L,I_R),d(p-d,p,I_R,I_L))$$
$$d(p,p-d,I_L,I_R)=\min_{p-d-0.5\le p-d+0.5}|I_L(p)-I_L(q)| \quad (3)$$

where $I(p)$ and $I(q)$ is the gray values of pixel p and q, respectively. And $d(p,p-d,I_L,I_R)$ represent the smallest difference between $I(q)$ and the three values of $I(p)$, $(I(p)+I(p-1))/2$, $(I(p)+I(p+1))/2$.

Because of the eight neighborhoods of a pixel, after calculating in one direction, it is necessary to expand to eight directions to calculate its minimum cost value separately. Then, the generative value in eight directions is accumulated, and the minimum accumulated cost value is selected as the final disparity value of the pixel. The images captured by binocular camera are all corrected images as shown in Fig. 2(a) and Fig. 2(b). The comparison of the results obtained by the algorithm in reference [3] and the algorithm in this paper is shown in Fig. 2(c) and Fig .2(d).



a. left image b. right image

c. ref. [3] d. algorithm of this paper

Fig. 2 Original and Contrast Graphs

By comparing the effect maps, we can see that the disparity image based on block matching has more noise points and worse image quality, while the image quality obtained by this method is better and fewer holes. Compared with the method used in reference [6], the time-consuming of this method is shorter. By experiments on the image with the size of 800*600, the time comparison effect of the three methods is shown in Table I. The data show that this method is more suitable for practical operation, and it takes less time and can obtain high quality images.

TABLE I
COMPARISONS OF RUNNING TIME OF DIFFERENT ALGORITHMS

Algorithm	Times(ms)
Ref. [3]	105
This paper's Algorithm	852
Ref. [6]	78112

III. image restoration

The image based on binocular stereo matching still has noise and holes, which affect the quality of the image. In this section, the disparity image obtained in the second section is filled with holes. There are four main reasons for the formation of holes: (1) the image taken by a binocular camera from one perspective can not be captured by a camera from another perspective, which is often set at 0 pixels; (2) during the process of camera shooting, the reflection of the object surface will also produce holes; (3) the object scene is beyond the scope of the camera; (4) Image After stereo matching, the pepper noise in the image itself will be retained in the disparity map to form a "holes". And image restoration method for holes filling and denoising are research hotspot at present. Criminisi algorithm is a classical filling algorithm based on the texture characteristics of image sample blocks. It establishes a patching template in the neighborhood for the pixels that need to be repaired. Then the priority of the patching template is determined. According to the priority order, the patching template is covered and filled by the regions which are close to the patching template. As shown in Fig. 3-1, $\partial\Omega$ is the repaired boundary and P is the pixel on the repaired boundary. I is the image to be repaired, Ω is the target area to be repaired, ψp is the patching template based on point P, and Φ is the area other than the target area.

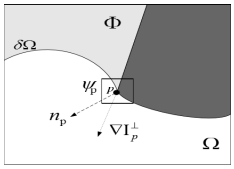


Fig. 3-1 Principle diagram of Criminisi algorithm

Taking the repaired pixels as the center, a rectangular template of suitable size is established, and the priority of each point in the template is calculated. The formula is as follows:

$$P(p)=C(p)D(p) \tag{4}$$

Among them, *C(p)* is confidence:

$$C(p)=\frac{n_p}{N_p} \tag{5}$$

n_p is the number of non-hollow points in the neighborhood around p-point, and N_p is the total number of pixels in the region.

D(p) is data item:

$$D(p)=\frac{|\nabla I_p^{\perp}\cdot n_p|}{\alpha} \tag{6}$$

Firstly, it calculates whether the pixels around P-point are valid points, and takes the direction of the maximum gradient as the gradient direction of the repaired area. n_p is the normal direction of the area, $\nabla I_p^{\perp}$ is the vector rotated 90 degrees in the gradient direction. α is the normalization factor. According to this formula, the points with more obvious structural features can be found.

The product of the two is a priority value. The more intact points in the template, the higher the confidence, that is, the higher the *C(p)*. The value of *D(p)* is related to the gradient value of the edge. After determining the priority level, the sample block closest to its texture feature is found by global matching for repairing. The judgment process uses the SSD algorithm, and the formula for calculating the SSD is:

$$d(\psi_p,\psi_q)=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}[(p_{ij}^R-q_{ij}^R)^2+(p_{ij}^G-q_{ij}^G)^2+(p_{ij}^B-q_{ij}^B)^2]} \tag{7}$$

Where *m* and *n* are the length and width of the area to be repaired, and *p* and *q* are the pixel values of the area and the candidate area to be repaired.

After the repairing is completed, the edge points of the repair are updated, and the above operations are repeated until the repair is completed.

The method proposed in this paper is roughly as follows:

(1) Since disparity images obtained by binocular stereo matching often have salt and pepper noise, the median filtering operation is first performed on such images. This paper selects a template size of 9*9 for images with a size of 800*600, which can better maintain edge information. And most independent noise points are effectively removed on the premise of keeping the image from being excessively smoothed. As shown in Fig. 3-2, the left image is the disparity map obtained by stereo matching, and the right image is processed by median filtering.

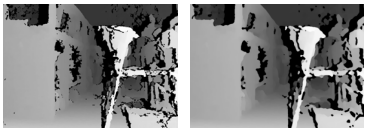


Fig. 3-2 Pre-filtering and filtering comparison

(2) On this basis, the point-by-point scanning of the image is performed: scanning from the middle of the image to the left and right. If multiple points (cavities) with zero pixels are scanned continuously, they are considered to be a more serious damage. As shown in Fig. 3-3, the inner area of the white rectangle is the severely damaged area. In view of this situation, the area with perfect points (the black rectangular box area) scanned at the previous time is judged, and the neighboring pixels around it are checked. If the variation is small, the pixel value is filled into the damaged area for repairing. For other cases where the hole is small (elliptical area), the neighborhood mean is used for repairing.

Fig. 3-3 Severe damage area

(3) Change the calculation formula for priority:

$$P(p)=C(p)+D(p)+G(p) \tag{8}$$

The original product form is changed to summation form, which changes the priority value, prevents the overall priority from being too small, reduces the cumulative error of repair, and introduces the gradient $G(p)$. Since there are unknown image information in the patch to be filled and it contains multiple image features, patches with higher gradient value are preferred to be repaired. Because the bigger the change of the pixel around the repaired points is, the bigger the gradient value of the repaired block is. So the priority of the repaired blocks over th flat areas is to achieve the purpose of enhancing the structural restoration performance.

(4) Find the damage point again, and judge the repair order according to the changed priority calculation formula. Take the 4*4 area around the damage point as the template to be repaired, and find other similar sample blocks on the image to fill. The flow chart of the algorithm in this paper is shown in Fig 3-4:

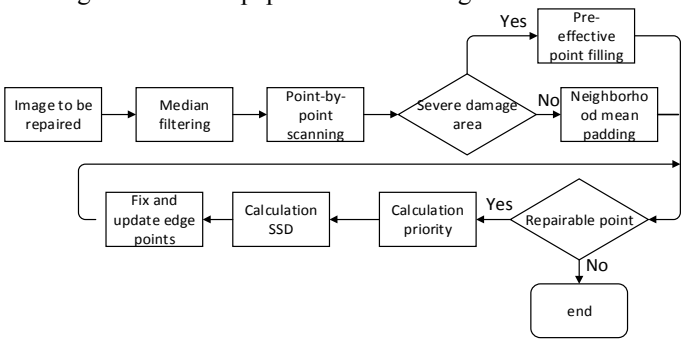


Fig. 3-4 Algorithm flow chart

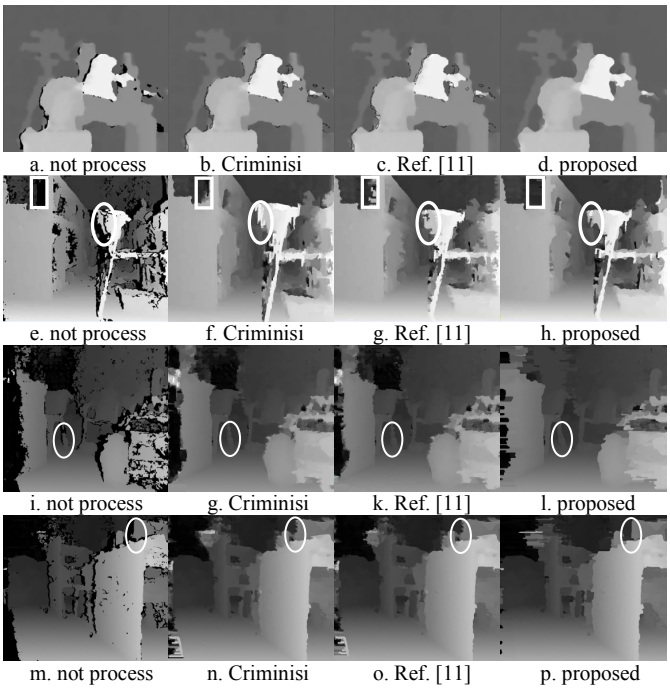
IV. experimental results

The experiment is operated based on a desktop PC. The processor is Intel (R) CoreTM i5-3470, and the operating system is 64-bit Windows 7 Professional. VS2012 is used as a compiler platform, and Opencv2. 4. 9 open source library is configured. The data sources used are standard images (face images) exposed on Middlebury platform and three real scene images captured in laboratory by Bumblebee II binocular camera. as shown in Fig. 4-1. They are face, chair, car, and front table.

a. face_left b. chair_left c. car_left d. table_left

Fig. 4-1 Image data

In this paper, the proposed algorithm is experimented on the acquired image and is compared with Criminisi algorithm and the algorithm proposed in reference [11]. The comparison results are shown in Fig. 4-2. From the first line to the last line, they are face, chair, car, and front table. Fig. 4-2(a), Fig. 4-2(e), Fig. 4-2(i), and Fig. 4-2(m) represent the unprocessed disparity map and the rest represent the repair results under different algorithms.



a. not process b. Criminisi c. Ref. [11] d. proposed
e. not process f. Criminisi g. Ref. [11] h. proposed
i. not process g. Criminisi k. Ref. [11] l. proposed
m. not process n. Criminisi o. Ref. [11] p. proposed

Fig. 4-2 Comparison of different algorithms

It can be seen from the information in the Fig. 4-2 that the Criminisi algorithm is prone to failure in repairing large-scale void areas (large black areas in the graph), and the restoration of regional structure is poor: the white ellipse selection in the graph. The method proposed in reference [11] saves a lot of time in processing, but the effect of preserving structural information is not ideal. And it cannot deal with large-scale damaged areas very well. The method proposed in this paper cannot only repair images, but also achieve good results in maintaining the structure of objects.

This paper objectively compares the processed images by introducing signal-to-noise ratio (SNR). By dividing the images into blocks, calculating the variance of each block, and then averaging the variance of all blocks. Finally, calculating the noise ratio of each block, and then calculating the signal-to-noise ratio by variance method. The bigger the SNR value, the smaller the image distortion which means the better the image quality. Table II compares the SNR values of each image under four algorithms. The method proposed in reference [12] is experimented on the image used in this paper, and the comparison is shown in the table.

TABLE II
IMAGE SIGNAL-TO-NOISE RATIO CONTRAST TABLE .

Image	Not process	Criminisi	Ref. [11]	Ref. [12]	Proposed
face	2.7769	2.9071	2.9147	2.9626	2.9774
chair	1.6448	2.5231	2.4431	2.5739	2.6328
car	1.2902	1.9275	2.0081	1.9420	2.0468
table	1.4424	1.8559	1.8629	1.8328	1.8642

From Table II, we can see that the SNR value of Criminisi algorithm and reference [11], reference [12] algorithm is lower than that of this algorithm by comparing the four images, which objectively proves that this algorithm can repair the image better.

The processing time of this method is also greatly shortened. The time contrast effect chart is shown in Fig. 4-3. Among them, the size of the standard face processing image is 480*360, and the other image size is 800*600. Large-scale damaged area and noise are the important reasons affecting the processing speed of Criminisi algorithm. In this paper, many "bad points" of the original image are pre-repaired, which reduces the repair time by about 30% compared with Criminisi algorithm. The time consumed by Ref. [11] algorithm is related to the selection of distance *n*. When there are fewer noise points and holes in the unprocessed image, the processing time of the method in reference [11] is faster. While for the severely damaged image, the processing time of the algorithm in this paper is faster.

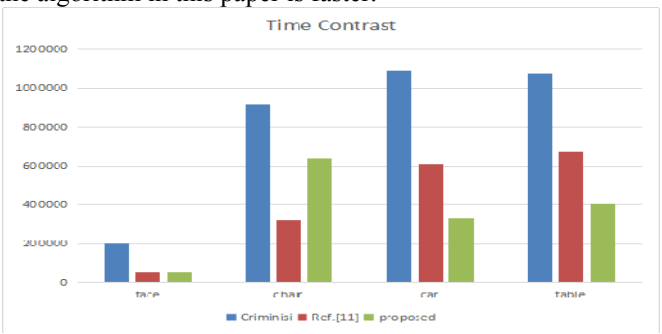


Fig. 4-3 Time comparison effect

V. conclusions

Aiming at the problem of obtaining and repairing disparity maps based on binocular cameras, this paper proposes a process from stereo matching to image restoration. Through this process, the images captured by binocular cameras can be transformed into disparity maps with high image quality. Firstly, a semi-global stereo matching algorithm is used to obtain the disparity map of the collected image. Aiming at the problem of image restoration, a sample block filling method combining median filtering and point-by-point scanning is proposed, and a gradient-first selection mechanism is introduced in the process of image restoration. This method can keep the edge structure information of the image better and improve the overall image quality greatly. The method still has some shortcomings, that is, the amount of calculation is still large and time-consuming, which is not conducive to real-time operation.

acknowledgment

This paper is supported by the Project for the National Natural Science Foundation of China under Grants No. 61672064 and the Beijing Natural Science Foundation under Grant No. 4172001.

references

[1] Chen Zhang, Lifeng Wang, Zhijun Meng. Binocular Depth Estimation Based on Diffractive Optical Elements and the Semiglobal Matching Algorithm[C] //Proceedings of IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), 2018: 27-29.

[2] Marr D, T poggio. Cooperative computation of stereo disparity[J]. Science, 1976, 194(4262): 283- 287.

[3] F Wang, K Jia, J Feng. The Real-Time Depth Map Obtainment Based on Stereo Matching[C] //Proceedings of Euro-china Conference on Intelligent Data Analysis and Applications. Berlin: Springer, 2016: 138-144.

[4] Roy S, Cox IJ. A maximum-flow formulation of the n-camera stereo correspondence problem[C] //Proceedings of IEEE International Conference on Computer Vision. India: Bombay, 1998: 492-499.

[5] Vieira G D S, Soares F A A M N, Laureano G T, et al. A Segmented Consistency Check Approach to Disparity Map Refinement[J]. Canadian Journal of Electrical and Computer Engineering, 2019, 41(4): 218-223.

[6] S Martull, M Peris, K Fukui. Realistic CG Stereo Image Dataset with Ground Truth Disparity Maps[J]. Technical Report of Ieice Prmu, 2012, 111: 117-118.

[7] Xinxin Yang, Jize Sun, Weimin Diao. Depth Image Inpainting for RGB-D Camera Based on Light Field EPI [C] //Proceedings of IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), 2018: 27-29.

[8] Fan Qian, Hu Xuelong, Zhang Lifeng. FMM algorithm based image in-painting method with an edge prediction[J]. Icic Express Letters, 2015, 9(5): 1419-1425.

[9] Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[10] Li Xufeng, Wang Jing, Liu Hongmin. Feature priority block matching algorithm for image restoration[J]. Journal of Computer-Aided Design and Computer Graphics, 2016, 28(7): 1131-1137.

[11] Li Aiju, Li Yujie, Niu Wenliang, Wang Tingmei. An improved criminisi algorithm-based image repair algorithm[C] // Proceedings of International Congress on Image and Signal Processing. USA: Institute of Electrical and Electronics Engineers Inc. 2016: 263-267.

[12] Liu J Z, Wu W H, Cheng C, et al. Depth image inpainting method based on pixel filtering and median filtering[J]. Journal of Optoelectronics•Laser, 2018, 29(5): 79-84.

[13] Lin Sen, Yin Xinyong, Tang Yandong. Research status and prospects of binocular visual stereo matching technology[J]. Science Technology and Engineering, 2017, 17(30): 135-147.