Proc. of the 9th Intl. Conf. of Asian Society for Precision Engg. and Nanotechnology (ASPEN 2022)
15–18 November 2022, Singapore. Edited by Nai Mui Ling Sharon and A. Senthil Kumar
doi://10.3850/978-981-18-6021-8_OR-02-0232

Accurate All-round 3D Measurement from Trinocular 360-degree Stereo Cameras via Geometric Optimization of Image Texture and Color

Takumi Hamada[1], Sarthak Pathak[1] and Kazunori Umeda[1]
1 Department of Precision Mechanical Engineering, Chuo University, 1-13-27, Kasuga Bunkyoku, Tokyo, 112-8551, Japan
Corresponding Author / Email: hamada@sensor.mech.chuo-u.jp, TEL: +81-3-3817-1826

KEYWORDS: Image processing, Spherical camera, 3D reconstruction, Stereo

Cameras are often used for Three-dimensional (3D) measurement in robotics and other applications. 3D reconstruction with cameras is useful in applications such as mapping at disaster sites that are inaccessible to humans and infrastructure inspections. Robots can move around in such locations and capture 3D data from multiple points of view. However, such data obtained from a moving robot needs to be integrated to obtain the reconstruction of the entire environment. This process is tedious, time consuming, and inaccurate. For this reason, we focus on a method that can accurately measure the environment in all directions at once.

Binocular stereo vision is one of the most common methods for 3D reconstruction. A stereo camera obtains distance information by using the disparity between images captured by two cameras at different viewpoints. With ordinary cameras, it is difficult to measure all directions at once because the field of view is limited. On the other hand, 360-degree cameras can capture all-round environment. However, it is practically difficult to achieve highly accurate all-round 3D measurement with binocular 360-degree stereo vision. This is because the accuracy in the epipolar direction, i.e., the direction along the two cameras, is extremely low.

In this research, we propose a method to increase the accuracy of all-round 3D Measurement using the principle of stereo cameras based on images obtained from Trinocular spherical stereo cameras. By placing a third camera at 90 degrees to two other cameras, it is possible to recover the loss of accuracy along epipolar directions. In this research, we aim to obtain high accuracy by using geometric and photometric constraints. We optimize each 3D point in the environment to satisfy the epipolar constraints of each camera pair, and to have the same intensity in each image. The accuracy is improved by weighting based on the confidence level and optimizing the reprojection error. We also account for the error in calculating stereo disparity by considering image gradient information in the optimization. In short, we reproject the measured points of the environment on all three cameras and minimize the reprojection error in a geometric manner, while considering the color and gradient information obtained from the images. Experimentally, we investigate the improvement in accuracy by three methods:
1. Geometric constraints - consistency of the projection of a 3D point in the environment onto three cameras
2. Color information - consistency of color information projected onto the three cameras
3. Reliability of disparity calculation by optical flow based on gradient information in the image

NOMENCLATURE

$d =$ distance
$w =$ weighting
$\epsilon =$ reprojection error
$\delta =$ color information

1. Introduction

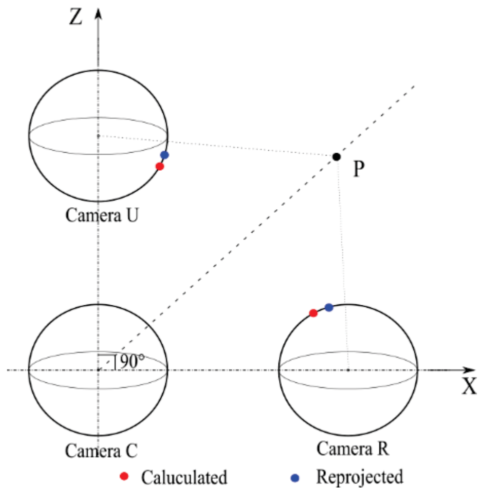
In recent years, cameras have played the role of the eyes of robots and are often used in research and applications. 3D measurement with cameras is very useful for automating robots and understanding terrain at disaster sites that are inaccessible to humans. One of the 3D reconstruction methods is binocular stereo [1]. A stereo camera obtains distance information by using the disparity. A 360-degree spherical camera can capture the all-round environment, which

ASPEN
Proc. of the 9th Intl. Conf. of Asian Society for Precision Engg. and Nanotechnology (ASPEN 2022)
15–18 November 2022, Singapore. Edited by Nai Mui Ling Sharon and A. Senthil Kumar

provides a large amount of information and enables measurement in a single shot. [2]. In previous research, when calculating pixel flow, the system is not able to handle cases where the texture has few features and cannot be calculated accurately. Therefore, this research proposes the use of the gradient and color information of the image. This research uses images captured by Trinocular spherical stereo cameras [3] to reconstruct the 3D shape of the environment and proposes a method to improve the accuracy by optimization using the gradient and color information of the images. In previous research, three spherical cameras were placed in an L-shape, and the three images taken were used as input images [4]. The two cameras were paired and the distance was calculated for each pair using the stereo camera principle. Stereo cameras have a problem that measurement accuracy decreases in the direction in which the cameras are aligned, and we improved the accuracy by weighting the cameras according to their angles and performing optimization. However, when calculating pixel flow, it is not possible to deal with cases where the texture has few features and cannot be calculated accurately. Therefore, in this research, we will improve the accuracy by capturing texture features from the gradient of the image and increasing the confidence level of pixels for which features are obtained. In addition, since it is expected that the colors of the corresponding pixels will match when reprojected, optimization was performed to eliminate the difference in color before and after reprojection. We have used these methods to improve accuracy.

2. Experimental Environment

Overview of the capture environment in this research is shown in Fig. 1.



(a) Camera Location

(b) Input Image at Camera U

(c) Input Image at Camera C

(d) Input Image at Camera R

Fig. 1 Camera Outline and Equitangular Images

The accuracy of the binocular spherical stereo system is reduced in the direction of the epipolar line, which is the direction in which the cameras are aligned, so an additional camera was added. Therefore, three cameras are arranged in an L-shape with a baseline length of 0.4 m.

3. Proposed method

3.1 Reprojection error and optimization

This research consists of four steps: distance measurement using the stereo camera principle, calculation of reprojection error, optimization of reprojection error, and reconstruction to a point cloud. Reprojection is the process of finding the corresponding pixel on the spherical camera from the 3D position obtained once. However, the corresponding pixel does not return to its original position due to reconstruction errors, and shift occurs. The difference between the reprojected point and the original point is the reprojection error.

The reprojection is performed in the following steps. First, three-dimensional coordinates are calculated from the distances obtained, and a three-dimensional reconstruction is performed with respect to camera C. Next, the reconstruction results are projected onto camera U and camera R shifted by the baseline length. In this case, the projection is made on the spherical image with a radius of 1. After projection, the reprojection error is obtained by taking the difference from the actual input image.

When the corresponding pixels of all cameras are projected in 3D, there is a constraint that the pixels must be in the same position and color. In addition, since the correspondence is determined by the optical flow [5], the points with the strongest gradients are more reliable. We calculate the disparity between the central and upper camera images, and the central and right camera images using

DeepFlow [6]. We developed an optimization method that takes these three phenomena into account. The Levenberg-Marquardt method was used for optimization. The distance from the reference camera was adjusted to minimize the reprojection error, and the distance at which the error was minimized was determined. Since there are two pairs of cameras in this method, the optimization is performed by taking the average of the distance information of the two obtained and using it as the initial value. The optimization is expressed in the following equation.

$$d_g = \arg min\left(w_{cu}\epsilon_{cu}(d) + w_{cr}\epsilon_{cr}(d)\right) \quad (1)$$
$$d_c = \arg min(\epsilon_{cu}(d) + \epsilon_{cr}(d) + \delta_c) \quad (2)$$

where d_g, d_c, $\epsilon_{cu}(d), \epsilon_{cr}(d)$ and w_{cu}, w_{cr} are the distance, reprojection error and weighting after optimization. δ_c is the color information.

3.2 Optimization based on image gradient information

Pixel flow, which is the amount of movement of corresponding pixels in two images, is calculated using optical flow. Optical flow calculates pixel flow by extracting the corresponding feature points in the image, but it does not work well for featureless uniform regions. Theoretically, there should be no pixel flow in the perpendicular to the epipolar direction. In actual measurement, however, lateral pixel flow occurs due to errors. To solve these problems, the image gradient information is used as a confidence level.

Specifically, we calculate the gradient in the epipolar direction and use it as a measure of the confidence of the disparity calculation by optical flow. The obtained edges are weighted and minimized using Equation (1) to obtain the optimal distance. The weights are visualized in grayscale in Fig. 2.

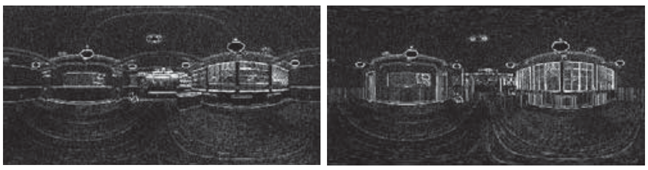


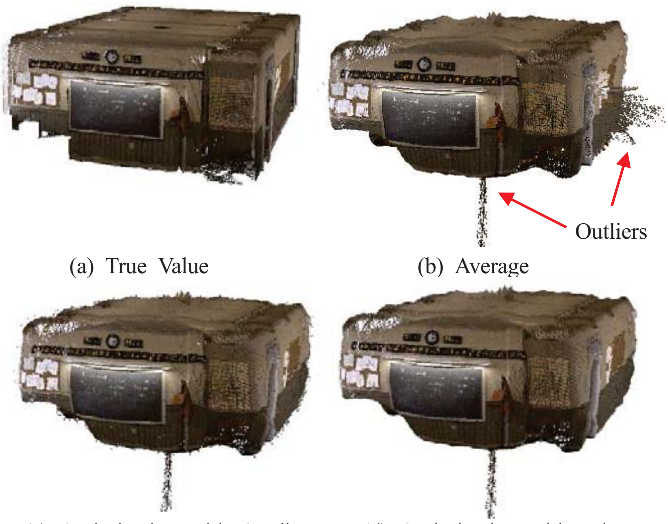
Fig. 2 Grayscale Images of Gradient

3.3 Optimization based on color information

We focused on the difference in color information per pixel before and after reprojection. The color information of the corresponding object is invariant before and after reprojection. Therefore, the difference in color information per pixel before and after reprojection is taken and optimized to minimize it using equation (2). In this research, a grayscale image was used to obtain consistent color information.

4. Results and Discussion

Fig. 3 shows the results of the reprojection using input images of resolution 1000 × 500 pixel. 3DCG software Blender was used to render an equitangular image in a virtual environment simulating a classroom as the input. The simulation environment provides the true values of the model, allowing quantitative evaluation.



(a) True Value (b) Average

(c) Optimization with Gradient (d) Optimization with color

Fig. 3 The Reconstruction Results

Fig. 3 (b) shows the reprojection result when the average of the measured distances is used without optimization. Fig. 3 (c) and (d) show the results of optimization using the gradient and color information of the image. The average absolute errors between the true value of the distances obtained in the virtual environment and the calculated distances illustrated in Fig. 4.

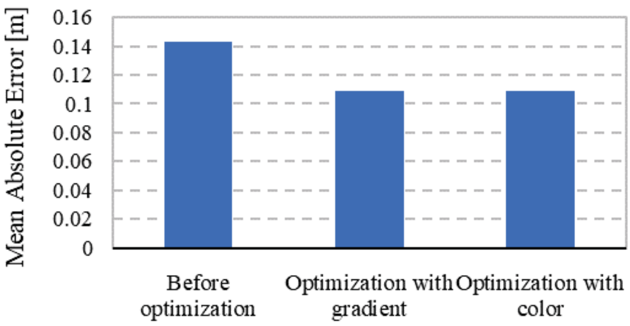


Fig. 4 Mean Absolute Error

In Fig. 3 (b), it can be seen that outliers occur as a property of the decrease in measurement accuracy in the direction of the epipolar line, which is a characteristic of stereo cameras. However, as shown in Fig.3 (c) and (d), these outliers were reduced by optimization using the gradient and color information of the image. As shown in Fig. 4, the mean absolute errors were 0.14337 m, 0.10906 m, and 0.10902 m before, after, and after optimization using the gradient and color information. The percentage reductions were 23.93% and 23.95% for the proposed optimizations using the image gradient information and the optimization using the color information, respectively. The optimization reduced the mean absolute error significantly from the pre-optimization value.

Fig. 5 Inside the Reconstruction classroom

However, as shown in Fig. 5, which was taken inside a classroom, the restored result could not completely correct the inaccurate areas. In addition, since the input image used in this research was a low-resolution image with a resolution of 1000 × 500 pixel, a highly accurate pixel flow could not be obtained, resulting in a rounded reprojection of the image as a whole. In order to improve these results, it is necessary to expand the color information from grayscale to RGB, and optimize the pixel correspondence from one pixel to another, taking into account the surrounding information.

5. Conclusions

This research proposes a measurement method that takes into account the gradient and color information of images to improve the accuracy of reconstruction using spherical stereo cameras. The proposed method successfully reduces outliers and the average absolute error by minimizing the reprojection error using the gradient and color information of the image, demonstrating the usefulness of the proposed method. In the future, we plan to use RGB values, optimize the method considering multiple pixels, and perform measurements in a real environment.

REFERENCES

1. S. Sakai, T. Takahashi, K. Ito, T. Aoki and H. Unten, “3D Reconstruction from Two Views Using Consumer Digital Camera,” Information Processing Society of Japan (JPSJ), vol.2011-CVIM-176 No.4 pp. 1-8, 2011.
2. Li, Shigang. "Trinocular spherical stereo." In 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4786-4791. IEEE, 2006.
3. S. Pathak, A. Moro, H. Fujii, A. Yamashita and H. Asama, “3D reconstruction of structures using spherical cameras with small motion,” 2016 16th International Conference on Control, Automation and Systems (ICCAS), October 2016.
4. W. Yin, S. Pathak, A. Moro, A. Yamashita and H. Asama, “Accurate All-round 3D Measurement Using Trinocular Spherical Stereovia Weighted Reprojection Error Minimization,” 2019 IEEE International Symposium on Multimedia (ISM), December 2019.
5. B. Horn and B. Schunck, “Determining optical flow,” Artificial Intelligence Volume 17, Issues 1–3, August 1981, Pages 185-203
6. P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, DeepFlow, “Large displacement optical flow with deep matching,” ICCV - IEEE International Conference on Computer Vision, Dec 2013, Sydney, Australia. pp.1385-1392.