Robustness of ToF and Stereo Fusion for High Accuracy Depth Maps

Zhen-shan BAO, Bo-wen LI and Wen-bo ZHANG*

Faculty of Information, Beijing University of Technology, No.100 Pingleyuan, Chaoyang District, Beijing, People’s Republic of China

*Corresponding author

Keywords: Depth map, Binocular stereo vision, Time-of-Flight, Fusion.

Abstract. Depth map can be used in many applications such as robotic navigation, driverless, video production and 3D reconstruction. Currently, both active depth cameras and passive stereo vision systems are the main technical means to obtain depth information, but each of the two systems alone has its own limitations. In this paper, the Time-of-Flight (ToF) and stereo fusion framework is proposed to solve the limitations of these two systems. The scheme in this paper contains "the prior fusion stage" and "the post fusion stage". Using depth information design energy function from ToF cameras to boost the stereo matching of passive stereo in "the prior fusion stage". During "the post fusion stage", the weighting function is designed according to the depth map of the stereo matching and the credibility map of the ToF depth map, and then the adaptive weighted depth fusion is performed. Experimental results clearly show that our method can obtain high-precision and high-resolution depth maps with better robustness than the other fusion approaches.

Introduction

For decades, the acquisition of deep information has attracted continuous attention from academia and industry. There are two ways to get the depth information of the scene, the passive depth acquisition and the active depth acquisition method. The passive depth acquisition method mainly performs stereo matching on the corresponding points of the image to obtain depth information. High-resolution depth maps can be obtained using the stereo matching algorithm, but it is difficult to calculate depth in untextured areas, repeating areas, and occlusion areas.

Active depth acquisition methods mainly include ToF depth camera or structured light depth camera, which can directly obtain the depth information of scenes. However, its resolution is much lower than that of a color camera. It is also susceptible to noise interference, and the error of obtaining depth information in a scene with severe scattering is extremely large. The depth camera can provide an initial depth reference to the 3D experience, which is not available in the stereo matching algorithm. Therefore, a system that combines a depth camera (active) and a binocular camera (passive) is currently a hot spot for obtaining depth information.

The algorithm proposed in this paper is based on the combination of depth information acquired by ToF and stereoscopic vision to achieve complementary advantages between them to obtain high-resolution, high-precision depth images.

Proposed Method

The algorithm proposed in this paper is mainly divided into three processes: Multisensor Calibration, "the prior fusion stage" and "the post fusion stage".

Multisensor Calibration

In the process of binocular stereo matching and parallax transformation we need to know the position relation between cameras, so as to calibrate the camera and correct the image. In particular, the distortion of ToF depth images is larger, and we need the internal parameters of ToF camera to be distorted and corrected. The specific process of calibration is to obtain checkerboard of different
positions through the imaging system. We obtain the parameters through the Zhang calibration algorithm [1].

The principle of stereo matching can be referred to [2]. We use the ToF-right camera as the system. Where $T_{rt}$ is the baseline length between the ToF camera and the right camera and $f$ is the ToF focal length. $H_{rt}$ is a homography matrix for ToF camera calibration in ToF-right camera system, it can be obtained by calibration. $Z$ is the depth value obtained by ToF, by formula (1) we can get the disparity map of ToF conversion.

$$D_{t}(t,r) = \frac{T_{rt} \ast f}{H_{rt} \ast Z}$$ (1)

Prior Fusion Stage

"The prior fusion stage" is to use the trusted depth information of ToF to guide the stereo matching algorithm to find more accurate matching points, so as to obtain the correct depth value.

The traditional stereo matching algorithm [3] is to find the corresponding matching point on the target view and calculates the parallax. The procedure to compute disparity usually requires three steps: local stereo matching, cost aggregation, and global optimization. The latter two steps are for the optimization and analysis of the initial local cost, which will be used in some algorithms. It is a well known fact that the local stereo matching is the most important part. The initial matching cost function is shown in equation (2).

$$C_{o}(x, y) = |I_{i}(x, y) - I_{r}(x_{0}, y_{0})|$$ (2)

Where $C_{o}(x, y)$ is the initial matching cost function, $I_{i}(x, y)$ and $I_{r}(x_{0}, y_{0})$ represent the coordinates of the matching point.

ToF can obtain an untrusted depth value in the low reflection area, such depth information obviously cannot guide the stereo matching to find the matching point. We can point out areas of unreliable depth through the intensity map of ToF modulation. The reason why ToF produces wrong depth information in these areas is that the light modulated by these objects returns to the depth camera is too small, making it impossible for the modulated light receiving and processing analysis module in the camera to measure the correct depth. In other words, in the ToF modulated light intensity map, the pixels whose light intensity value is less than a threshold value, the depth of the pixel point in the corresponding depth map can be considered as unreliable. Conversely, if the modulation intensity of a certain pixel is stronger, the more stable and credible the depth measurement is[4]. We make the intensity map of the ToF modulation A. For each pixel $(x_{0}, y_{0})$, it corresponds to the pixel of the ToF depth map. We define the trusted weight $W_{A}(x_{0}, y_{0})$ of the intensity map as show in equation(3).

$$W_{A}(x_{0}, y_{0}) = \begin{cases} \log A(x_{0}, y_{0}) + \delta, & \gamma \\ 0, & \text{otherwise} \end{cases}$$ (3)

where $\gamma$ and $\delta$ are manually predefined parameters.

In order to compensate for the damage to the accurate area of the ToF depth measurement using the global optimization algorithm, we design the energy function difference penalty term of the depth map. The penalty item $C_{p}$ is defined as show in equation(4).

$$C_{p}(x, y) = e^{k_{s} - P_{s}(x_{0}, y_{0})/\theta}$$ (4)

Where $P_{s}(x_{0}, y_{0})$ represents the value of the ToF measurement depth of the pixel converted to the parallax, $\theta$ is a threshold selected based on experience and used to control the value range of the penalty amount $C_{p}$. Through this penalty design, the matching range of stereo matching in the repeated texture region and the low texture region will be determined according to the trusted ToF
depth value, which can effectively improve the matching effect. According to the above description, the energy cost function of stereo matching is defined as show in equation (5).

\[ C_D = W_s(x_0, y_0) \times C_f(x, y_0) \]  

(5)

After defining the depth energy function of the ToF, the local stereo matching process can be assisted by the reliable area of the ToF. Finally, the total stereo matching cost function is defined as show in equation (6).

\[ C(x, y_0) = C_o(x, y_0) + C_D(x, y_0) \]  

(6)

Then we aggregate according to the corresponding cost, and get the optimized stereo matching disparity map \( D_s(x_0, y_0) \). And here we have completed "the prior fusion stage" part.

**Post Fusion Stage**

After completing "the prior fusion stage", we have obtained an optimized stereo matching disparity map \( D_s(x_0, y_0) \). Using ToF reliable depth information to guide stereo matching, the errors in weak texture region and deep discontinuous area are reduced, but still no good improvement in the occlusion area. According to equations (7) and (11), we make the confidence maps of the optimized disparity maps \( D_s(x_0, y_0) \) and ToF disparity maps. Designing their corresponding weight functions for untrusted regions. Where \( r_s \) is the reliability of the disparity map, \( C_{1st} \) is the minimum matching cost obtained in the local stereo matching algorithm, and \( C_{2nd} \) is the second matching cost obtained. we believe it is correct for the pixel disparity value with a reliability greater than 0.1.

\[ r_s = 1 - \frac{C_{1st}}{C_{2nd}} \]  

(7)

For pixel parallax with a confidence of less than 0.1, we use the parallax data weighting method for fusion. That is, the disparity map obtained by the ToF is \( D_t(x_0, y_0) \), the disparity value of this part can be considered to be mainly caused by the occlusion area. This partial disparity \( D \) as show in equation (8).

\[ D = W_s \times D_s + W_t \times D_t \]  

(8)

\[ W_s = \frac{r_s}{r_s + r_t} \]  

(9)

\[ W_t = 1 - W_s \]  

(10)

\( r_s \) is the degree of confidence that the disparity is worth for the ToF. According to the experimental results in [5]. The reliability of the disparity map by ToF as show in equation (11).

\[ r_t = \frac{1}{A(p)} \]  

(11)

Where \( A(p) \) is the standard deviation of the ToF amplitude diagram.

In this way, we design the ToF disparity map and the weighted function of the stereo matching disparity map based on the credibility map, and carry out an adaptive weighted fusion method to make up for the error of the stereo matching in the occlusion area. Afterwards, the parallax map is converted into a depth map according to the calibration results of the fourth section, and a high-resolution, high-precision depth image is finally obtained.

**Experiments**

This paper uses the fusion approach data set provided by the Multimedia Technology and Telecommunications Lab of the University of Padova to verify the correctness of the proposed algorithm [6]. We use the approach proposed in the paper to fuse different scenes. The results
obtained are compared with the standard depth maps. By comparing with other fusion approaches, we verify the advantages of the proposed approach.

In order to show the experimental results of the algorithm, we provide several different scenarios, including weak texture regions, occlusion regions, and heavily scattered regions. The experimental scene is shown in Figure 1. Stereo matching algorithm we use the traditional stereo matching algorithm based on adaptive window selection [7]. We also compare the data with other fusion algorithms. Here we use the root mean square (RMSE) error compared with the standard depth map to show the performance of the proposed algorithm.

![Figure 1. Comparison of experimental results.](image)

We used the average of RMSE in 20 scenarios in the dataset to evaluate the advantages of the proposed method in fusion accuracy. Table 1 gives the root mean square error (RMSE) of the final disparity map. It can be seen that our method can significantly improve the depth map quality of ToF and the quality of depth maps using stereo matching alone, improve the precision and resolution of depth map, and have advantages in different scenes, and verify the robustness of the algorithm.

| Resolution   | RMSE (m) |
|--------------|----------|
| Original ToF map | 424*512 | 0.0871 |
| Stereo matching[7] | 540*960 | 0.0772 |
| "prior stage"   | 540*960 | 0.0458 |
| "post stage"    | 540*960 | 0.0372 |

Table 2 shows that our method has better real-time and robustness compared with other algorithms.
Table 2. The performance of this method is compared with other methods.

| Method                  | RMSE (m) | Time (ms) |
|------------------------|----------|-----------|
| Q. Yang [8]            | 0.0739   | 185       |
| J. Zhu [9]             | 0.0432   | 112       |
| Method of this paper   | 0.0372   | 95        |

**Conclusions**

In this paper, we propose a robust approach to obtain high accuracy depth maps by the integration of passive stereo and ToF cameras. The main contributions of this paper is to use the depth information of ToF to assist stereo vision systems and obtain a more accurate depth information through an approach about "the prior fusion stage" and "the post fusion stage". Experiments show that the proposed approach achieves the improved results with high accuracy and robustness.

In the future, we hope that there will be some improvement in the speed of the fusion approach. We will consider adopting new high-performance computing devices to enhance efficiency.

**References**

[1] Z. Zhang, A Flexible New Technique for Camera Calibration, J. IEEE Transactions On Pattern Analysis and Machine Intelligence (TPAMI), 2000, 1330-1334.

[2] MZ. Brown. Advances in Computational Stereo, J. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 993-1008.

[3] J. Cheng, C. Leng. Fast and Accurate Image Matching with Cascade Hashing for 3D Reconstruction, C. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, 23-28.

[4] S. Hussmann. One-Phase Algorithm for Continuous Wave TOF Machine Vision Applications, C. Proceedings of the IEEE Transactions on Instrumentation and Measurement, 2013, 991-998.

[5] SR. Guomundsson, H. Aans and R. Larsen. Environmental effects on measurement uncertainties of time-of-flight cameras, J. Mathematics of Computation, 2007, 30(30):143-143.

[6] G. Agresti, L. Minto. Deep Learning for Confidence Information in Stereo and ToF Data Fusion, C. Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW), 2017, 697-705.

[7] J. Zhang, DX. Li and M. Zhang. Fast stereo matching algorithm based on adaptive window, J. International Conference on Audio Language & Image Processing, 2010, 16(6), 138-142.

[8] Q. Yang, KH. Tan and B. Culbertson. Fusion of Active and Passive Sensor for Fast 3D Capture, C. Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP), 2010, 69-74.

[9] J. Zhu, L. Wang. Reliability Fusion of Time-of-Flight Depth and Stereo for High Quality Depth Maps, J. IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI), 2011, 33(7):1400-1414.