Three-dimensional reconstruction method based on jitter optimization

Peichen Liang, Weixin Zheng, Xiao Xu*

School of Physics and Optoelectronics, South China University of Technology, Guangzhou, Guangdong 510640, China

*Email: physicsxuxiao@163.com

Abstract. For obtaining richer visual information and eliminating mismatched points faster, a jitter-optimized 3D reconstruction algorithm is proposed. First, we built a stereo vision system that can capture images from multiple views of a jittering plane mirror. A vector-based method is defined to evaluate the performance of 3D point cloud and predict the optimal jittering state of the object, and then select robust feature points with integration optimization strategy. Experimental results show that the proposed algorithm can efficiently eliminate mismatched points and supplement robust features. Furthermore, selecting a reasonable visual angle of the binocular cameras will significantly improve the effect of jitter optimization, and higher-quality reconstruction will be obtained. The proposed method can solve the difficult problem to change the pose of the object or adjust the position of the binocular cameras to capture images from multiple views.

Keywords. stereo vision; 3D reconstruction; state vector; jitter optimization

1. Introduction
Stereo matching is the most important and challenging step for the process of binocular 3D reconstruction. Stereo matching algorithm mainly includes matching cost computation, cost aggregation, disparity calculation and disparity refinement [1-3]. In order to improve the performance of 3D reconstruction, a RANSAC algorithm is usually applied to eliminate mismatched points in the process of stereo matching [4-5]. In addition, point cloud filtering which includes distance-based method [6] and density-based method [7] is also an effective approach to obtain more robust features.

A trinocular vision method [8] of 3D reconstruction, which captures images from multiple views, has been proposed to obtain richer features and improve the matching accuracy. However, it is difficult to control multiple cameras and ensure their synchronization and stability in actual scene when capturing images. Yang et al. [9] move a monocular camera to capture images from multiple views in different positions when implementing 3D reconstruction of cucumber leaf morphology. This approach may cause the camera’s parameters to change as the camera moves. Li et al. [10] proposes a jittering algorithm to select features, changing the position of the binocular cameras and implementing 3D reconstruction respectively. Then, two groups of point clouds are utilized to remove unstable features by comparing their difference. However, because of the non-ideal conditions in the real systems and the unoptimized jittering direction of cameras, the performance of the method is not good enough in real application scene.

In response to the above problems, we build a binocular 3D reconstruction system that can simulate camera jitter based on the method proposed by Xu et al. [11]. The target object can be captured at
different positions or views through a plane mirror without moving the cameras, which avoiding changes in the camera parameters and ensuring the stability of the stereo vision system. With the help of the jitter of the plane mirror, we can flexibly capture images from multiple views for 3D reconstruction. Moreover, a jitter optimization algorithm is proposed, which can reasonably predict the optimal jittering direction of the object and select robust features efficiently.

2. Jitter experiment system
Figure 1 shows the proposed experimental system of stereo vision. The plane mirror controlled by a motor is placed in the field of view of the binocular cameras, which can perform translation, rotation and pitching motions at a series of specified amplitudes respectively. In the experiment, the target object is placed between the two cameras. In fact, the position can be flexible, provided that the reflected image of the object in the plane mirror can be captured.

By setting the motor’s jittering parameters $\phi = (t, \omega, \varphi)$, where $t$, $\omega$, and $\varphi$ denote translation, rotation and pitching respectively, the relative position of the object can be changed through plane mirror, and the information of captured images will be different. Thus, the process can be equivalent to capturing images and implementing 3D reconstruction of a fixed object from different positions or different views.

3. Proposed algorithm

3.1. The model of jitter optimization algorithm
The purpose of jitter is to obtain multi-view image information for correction of the 3D reconstruction system. However, single jitter operation can only supplement limited information, while scan-based jitter is time-consuming and computationally expensive. To solve this problem, a jitter optimization algorithm, shown in Figure 2, is proposed.

In order to facilitate the description of the jittering state, we use the notation $D^0$ to represent the point cloud reconstructed before the jitter, then $D^n (n = 1, 2, 3, ...) $ denotes that of the $n^{th}$ jitter operation, also known as the $n^{th}$ jittering state. In addition, we introduce the concept of jittering state vector (JSV), which is used to describe the features of each jittering state. These features reflect the quality of the reconstruction in the current state. In the experiment, we calculate the vector features of each jittering state, compare them between adjacent jittering states, and then employ the integration optimization strategy to refine the point cloud.
3.2. Calculate the jittering state vector
A method for evaluating the jittering state vector of single point cloud is proposed. The method considers that the mismatched points can reflect the principal error features of the point cloud in current jittering state and the direction of the next jitter operation can be reasonably predicted by calculating the jittering state vector features which includes the magnitude and the direction of error of mismatched points.

Table 1 shows the method of calculating the JSV. We first calculate the distance between each point and its neighborhood in the point cloud, and determine the magnitude and direction of the error of each point to select mismatched points. Assuming that the adjacent plane $N$ of a certain point in space is defined as the plane determined by the three nearest points $(P_1, P_2, P_3)$, then the error direction $V_p$ can be denoted as the opposite direction from the point $P$ to the plane $N$, and the error magnitude $E_p$ is measured by distance-weighted method in formula (1), where $\text{Dist}_N$ is the distance from point $P$ to plane $N$, $\text{Dist}_P = \sum_{k=1}^{3}(P - P_k)^2$ is the sum of the distances from point $P$ to the three points that make up the plane $N$, $w_s \in [0,1]$ is the weight of distance.

$$E_p = w_s \cdot \text{Dist}_N + (1 - w_s) \cdot \text{Dist}_P$$

The point whose error $E_p$ is greater than the given distance threshold $T_s$ is defined as a mismatched point which will be added to point set $M$. Then, we define the magnitude of jittering state error $E_S$ as the sum of the errors of all points in $M$, that is, $E_S = \sum_{P \in M} E_p$, and the direction of the jittering state vector $V_S$ is defined as the principal direction of mismatched points’ error vector.

**Table 1. Calculation of jittering state vector.**

| I. | Initialize $M = \emptyset$, for each point $P$ in the point cloud |
|---|---|
| (i) | Calculate the error magnitude of $P$, that is, $E_p = w_s \cdot \text{Dist}_N + (1 - w_s) \cdot \text{Dist}_P$ |
| (a) | If $E_p > T_s$, then $P$ is a mismatched point, add $P$ to $M$ |
| (b) | If $E_p \leq T_s$, then $P$ is not a mismatched point |
| (ii) | Calculate the error direction $V_p$ of $P$ |
| II. | Count and analyze the error magnitude $E_p$ and direction $V_p$ of each point in $M$ |
| (i) | Calculate the total magnitude of the jittering state error: $E_S = \sum_{P \in M} E_p$ |
| (ii) | Calculate the direction of the jittering state vector: $V_S$ |
| (a) | Divide $0-360^\circ$ into 12 histogram bins, and allocate the direction of each point |
| (b) | $V_S$ is given by the direction of the bin with the most mismatched points |

The JSV, including the error magnitude $E_S$ and the direction $V_S$, reflects the statistical characteristics of the point cloud of the current jittering state, and provides an effective basis for further improving the effect of 3D reconstruction. Obviously, the object should move in the direction in which the total error is reduced, that is, the opposite direction of the JSV.

3.3. Adjacent state vector comparison
In order to obtain more robust vector features, multiple jittering states in the past time should be taken into consideration and used as a criterion to predict a new jittering vector. We define the adjacent state vector (ASV) to represent the result of comparing the current jittering state and past jittering states. As shown in Table 2, a “two out of three” strategy is adopted to calculate the ASV. The strategy considers the relative errors between three adjacent jittering states, and the jittering states with the maximum and minimum error will determine the direction of the ASV, while the two states with smaller errors will be used in the next jitter operation. In formula (2), $E_A$ is the error of the current adjacent state, $E^n_{a-2}$, $E^n_{a-1}$ and $E^n_a$ are the relative errors of the three jittering states in the “two out of three” operation, and $V_A$ is the direction of current adjacent state vector which is defined as the direction where the maximum error points to the minimum error.

$$\begin{align*}
E_A &= \min(E^n_{a-2}, E^n_{a-1}, E^n_a) \\
V_A &= V_{E_{\max}^a \rightarrow E_{\min}^a}
\end{align*}$$
Combining the JSV with ASV, we will obtain a more accurate jittering direction $V_T$. As shown in formula (3), $V_T$ is weighted by JSV and ASV, $V_{S_{\text{min}}}$ is the direction of JSV with the minimum relative error in the “two out of three” operation, and $\alpha$ and $\beta$ are the weights of direction vector.

$$V_T = \alpha \cdot V_{S_{\text{min}}} + \beta \cdot V_A$$

(3)

Table 2. “Two out of three” strategy.

| Step | Description |
|------|-------------|
| I.   | given the jittering threshold $\Delta\Phi = (\Delta t, \Delta \omega, \Delta \phi)$, jitter randomly within the threshold to generate three jittering states denoted as $D^1, D^2, D^3$, and abbreviated as $D^{1-3}$ |
| II.  | implement “two out of three” operation on state $D^{1-3}$: |
| (i)  | calculate the relative errors $E_{a_{\text{1-3}}}$ of $D^{1-3}$, where the jittering states with the maximum and minimum error will determine the direction of the ASV, that is $V_A$ |
| (ii) | predict: a more accurate jittering direction $V_T$ is jointly predicted by $V_A$ and $V_{S_{\text{min}}}$ |
| (iii)| jitter operation: start from the state with minimum error, and jitter along the direction $V_T$ to obtain a new jittering state, denoted as $D^4$ |
| (iv) | update: replace the state with the maximum error in $D^{1-3}$ with the predicted $D^4$ |
| III. | according to the updated states, repeat step II to continuously predict the new jittering state $D^5, D^6, D^7, ...$ |

3.4. Integration optimization strategy

In Table 3, an integration optimization strategy (IOS) is utilized to refine the point cloud. According to the previous analysis, the fastest improvement direction of error can be predicted. Thus, along the predicted direction, integral processing will perform better. We define several adjacent jittering states as an integral interval, and the number of jittering states in the integral interval is called integral stride. Integral processing will follow these two strategies:

- a) Points that remain robust in each jittering state within integral interval will be added to the reference point set $B$
- b) Constantly adding new robust points to $B$ during the jitter optimization process

Integral processing carries out after each jitter operation and a robust point cloud will be obtained when iteration to integral error convergence.

Table 3. Integration optimization strategy.

| Step | Description |
|------|-------------|
| I.   | set parameters: integral stride $s = 5$, initialize reference point set $B = \emptyset$ |
| II.  | if the current jittering state is $D^n$, then the current integral error $E^n_\text{int}$ is determined by the five jittering states in integral interval $[D^{n-s+1}, D^n]$, that is $D^{n-4} \sim D^n$ |
| III. | count the robust point set $C$ in $D^{n-4} \sim D^n$:
| (i)  | set the neighborhood window size $W$ |
| (ii) | take the point cloud $D^{n-2}$ as the benchmark, traverse each point $P$ in $D^{n-2}$:
| (a)  | select the point close to $P$ in $D^{n-4}, D^{n-3}, D^{n-1}, D^n$ within the neighborhood window |
| (b)  | if exist, save the similar points above in $D^{n-4} \sim D^n$, otherwise $P$ is not a robust point |
| (iii)| average the similar points as the robust point set $C$ when the traversal is over |
| IV.  | calculate $E^n_\text{int}$ by comparing the point set $C$ with the point set $B$ |
| (i)  | traverse each point $P$ in $C$ and calculate the distance from $P$ to the closest point in $B$ |
| (ii) | accumulate the distance of each point: $E^n_\text{int} = \sum_{P \in C} \text{dist}(B, P)$ |
| V.   | readjust the integral interval according to the new jittering state, repeat steps III and IV until the error converges |

4. Experiment

4.1. Experimental setup
A David's plaster statue is used as the experimental object for 3D reconstruction. To evaluate the accuracy of the proposed algorithm, we first reconstruct the point cloud model based on line laser scanning in Figure 3.

In addition, by changing the deflection angles of the two cameras, we set up four different experiments to explore the effect of jitter optimization under different visual angles. The parameter settings are shown in Table 4, where $\alpha_L$ and $\alpha_R$ is the angle between the optical axis of the left and right camera and the baseline, respectively, $\theta$ is the angle between the optical axis of the left and right cameras, that is, the visual angle.

| Table 4. Binocular camera deflection angle setting. |
|:----------------:|:----------------:|:----------------:|:----------------:|
| $\alpha_L/^{\circ}$ | 88.9927 | 88.1527 | 85.1372 | 83.0312 |
| $\alpha_R/^{\circ}$ | 89.8912 | 88.8890 | 89.9092 | 79.8801 |
| $\theta/^{\circ}$ | 1.1161 | 2.9583 | 4.9536 | 17.0886 |

4.2. The influence of integral stride
The integration optimization process is based on the information of the past and current jitting states, so different integral strides may bring different effects. Thus, we set up 10 different integral strides, and conducted 10 jitter optimization tests for each stride. Figure 4 shows the average errors of 10 jitter optimization tests under each integral stride. The curves with different colors and marks are four groups of jitter optimization experiments under different visual angles. The dashed and solid lines are the average errors of point cloud comparison with the laser scanning point cloud model before and after jitter optimization, respectively.
Figure 4. The influence of different integral strides.

In Figure 4, the average errors reduce gradually with the increase of the integral stride, and are all lower than initial errors before jitter, which proves that the integration optimization strategy can effectively obtain robust feature points and improve the accuracy of reconstruction.

4.3. Performance comparison

In order to further illustrate the effect of the proposed algorithm, we analyze the performance of 3D reconstruction when the visual angles are $\theta_1 = 1.1161^\circ$, $\theta_2 = 2.9583^\circ$, $\theta_3 = 4.9536^\circ$ and $\theta_4 = 17.0886^\circ$, respectively. As is shown in Figure 4, the initial error increases with the visual angle increasing ($\theta_1 \rightarrow \theta_4$). This is because the parallax between the two cameras is gradually expanding, and the difficulty of stereo matching increases. Thus, without jitter optimization processing, a smaller visual angle will obtain a better accuracy of stereo matching while the larger has worse.

Moreover, visual angles affect the performance of proposed algorithm. When the visual angle is small, stereo matching perform well, the improvement of the jitter optimization with small integral stride is unobvious, so the error of $\theta_2 = 2.9583^\circ$ is greater than the error of $\theta_1 = 1.1161^\circ$. With the increasing of integral stride, the effect of the jitter optimization will be dominant, so that the error of $\theta_2 = 2.9583^\circ$ becomes lower. When the visual angle is 4.9536°, the reasonable parallax significantly improves the effect of jitter optimization, and even a small integral stride can reduce the error greatly. However, for the visual angle of $\theta_4 = 17.0886^\circ$, due to the large difference between the left and right images, the stereo matching will be very difficult. Even though the error can be greatly improved through proposed algorithm, the effect is still unacceptable.

Figure 5 and Figure 6 respectively show the comparison of the effect of 3D reconstruction without optimization processing and using the proposed algorithm under four different visual angles. The first row in the figure is the reconstructed point cloud, and the second row is the result after surface fitting of the point cloud. Obviously, the reconstruction without jitter optimization has poor performance due to the existence of some outliers, and the effect of surface fitting is very unsatisfactory. After the application of the proposed algorithm, the reconstructed point clouds perform well because some outliers were removed and more robust features were added.
5. Conclusion
In this paper, we propose a jitter optimization algorithm for 3D reconstruction, which can dynamically establish the state vector of the object and predict the optimal direction of eliminating the mismatched points. In order to verify the algorithm in this paper, we conducted jitter operation with a plane mirror, and carried out several groups of comparative experiments by setting different visual angles and integral strides. The experimental results show that the proposed algorithm can improve the effect of reconstruction and the effect becomes more significant with the increasing of integral stride. However,
the effect is improved unobviously with a small visual angle and is significantly reduced with too large visual angle due to the greatly increasing difficulty of stereo matching. Thus, reasonable selection of visual angle of the binocular cameras can better exert the effect of the jitter method and improve the accuracy of 3D reconstruction. Moreover, the proposed method can solve the difficult problem to change the pose of the object or adjust the position of the binocular cameras to capture from multiple views.

6. Reference

[1] Zhang K, Fang Y and Min D et al. 2014 Cross-scale cost aggregation for stereo matching 2014 IEEE Conference on Computer Vision and Pattern Recognition (Columbus: IEEE) pp 1590–97

[2] Wang W and Zhang C 2012 Local disparity refinement with disparity inheritance 2012 Symp. on Photonics and Optoelectronics (Shanghai) pp 1–4

[3] Xiao J, Cheng H, Sawhney H, Rao C and Isnardi M 2006 Bilateral filtering-based optical flow estimation with occlusion detection European Conference on Computer Vision (vol 3951) ed Leonardis A et al. (Berlin: Springer) pp 211–224

[4] Chen W and Cao Q 2018 Feature points extraction and matching based on improved surf algorithm 2018 IEEE Int. Conf. on Mechatronics and Automation(ICMA) (Changchun: IEEE) pp 1194–98

[5] Liu J and Qiu Y 2011 Application of SIFT feature extraction algorithm on the image registration IEEE 2011 10th International Conference on Electronic Measurement Instruments (vol 3) (Chengdu: IEEE) pp 177–180

[6] Sotoodeh S 2007 Hierarchical clustered outlier detection in laser scanner point clouds Int. Arch. Photogrammetry, Remote Sensing and Spatial Information Sciences (vol XXXVI) ed Rönnholm P et al. pp 383–388

[7] Zaman F, Wong Y P and Ng B Y 2017 Density-based denoising of point cloud 9th Int. Conf. on Robotic, Vision, Signal Processing and Power Applications (vol 398) ed Ibrahim H et al. (Singapore: Springer) pp 287–295

[8] Shao M and Dong J 2019 Trinocular vision system based on cameras with tilt–shift lens J. Journal of Optics 48 43–48

[9] Yang L, Guo X, Liu X and Zhao C 2009 3D morphological reconstruction of cucumber leaf based on multiple images J. Transactions of the CSAE 25 pp 141–144

[10] Li R and Xu X 2019 Simulation of feature points screening by jitter method J. Transducer and Microsystem Technologies 38(05) pp 30–33

[11] Xu X and Xu S 2018 Binocular 3D reconstruction method based on jitter CN109636903A (2018) (Chinese Patent)