Fast Foot Measurement System Based on 3D Measurement Network

Qijian TANG[1], Xinyun LIU[1], Chang LIU[1], Xiaoli LIU[1], Menglong LIU[2], Xiang PENG*[1,2]
[1] Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province, College of Optoelectronic Engineering, Shenzhen University, Shenzhen Guangdong, China;
[2] Shenzhen ESUN Co. Ltd., Shenzhen Guangdong, China

Abstract

A multi-node 3D foot measurement network system is proposed based on digital speckle correlation method. For each node, two cameras build a binocular system, the corresponding points between the left and right cameras are established through the spatial correlation method. In order to improve the search speed of corresponding points, the epipolar and parallax constraints are adopted. The sub-pixel corresponding point location is achieved by Newton-Raphson iterative optimization method to improve the reconstruction accuracy, and the single-node depth data is finally reconstructed based on the stereo vision principle. Multi-view depth data is matched by means of the measurement network calibration technology. The whole foot data acquisition takes less than one second, which has a great commercial application prospect.

Keywords: 3d foot measurement, multi-node, speckle correlation, measurement network

1. Introduction

With the development of modern manufacturing technology, it will be a common phenomenon to obtain 3D data of human body for private customization. The human foot data can be widely used in the fields of personalized shoe customization, medical foot correction, shoe design and processing, ergonomic foot statistics, and so on. For foot measurement, instead of traditional manual measurement, some optical methods such as the laser scanning, structured light measurement are proposed [1]. Although the structured light 3D measurement technology has the advantages of non-contact, low cost and high accuracy and it has broad application prospects in industrial detection, intelligent manufacturing and other fields [3], the scanning time for obtaining the whole foot model is about 30 seconds or more at present, which is unacceptable for the old and the children, as customers should keep stationary during the measurement process. In this paper, a multi-node 3D foot measurement network system based on digital correlation method is proposed, which is realized by projecting speckle pattern onto the surface of object and simultaneously capturing a pair of random speckle images from the left and right cameras. With multiple viewpoints, the whole foot data acquisition takes less than one second.

2. Measurement principle

2.1 Binocular stereo vision

For a common binocular measurement system, as shown in Figure 1, O_l and O_r are the optical centers of the left and right cameras. The coordinate of point M is X_w in the world coordinate system, and is X_l in the left camera reference frame. Its image projections in the left and right image coordinate systems are denoted by m_l and m_r, respectively. $[R_s|T_s]$ is the position and orientation transformation matrix from the left camera to the right camera, and $[R_l|T_l]$ is the transformation from the world frame to the left camera. Generally, to simplify the model, the left camera reference frame is often chosen as the world frame, as a result, $R_l = I, T_l = 0$, where I is an identity matrix.
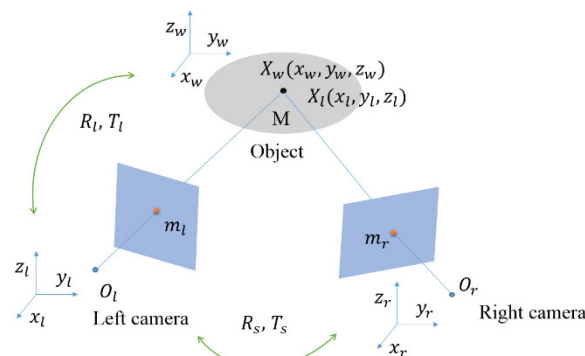

Fig. 1. Binocular stereo vision principle

*   xpeng@szu.edu.cn

According to the principle of line of sight intersection, the coordinate of point M can be calculated by:

$$\begin{cases} s_l \hat{m}_l = K_l [I|0] X_l \\ s_r \hat{m}_r = K_r [R_s|T_s] X_l \\ K_l = \begin{bmatrix} f_{xl} & \alpha_l & u_{0l} \\ 0 & f_{yl} & v_{0l} \\ 0 & 0 & 1 \end{bmatrix} \\ K_r = \begin{bmatrix} f_{xr} & \alpha_r & u_{0r} \\ 0 & f_{yr} & v_{0r} \\ 0 & 0 & 1 \end{bmatrix} \end{cases} \tag{1}$$

where K_i (i denotes l and r) is the intrinsic camera parameter matrix; f_{xi} and f_{yi} are the equivalent focal length; (u_{0i}, v_{0i}) is the principal point coordinate; α_i is the skew coefficient defined by the angle between the u and v pixel axes; $\hat{m}_i$ is image coordinate after lens distortion correction based on Brown-Conrady model; s_i is the scalar factor. These parameters are easily acquired in the system calibration process [3].

2.2 Correlation model

Taking the point M in the left image in Figure 2 as an example, the correlation region size is $(2w_m + 1) \times (2w_m + 1)$ pixels. The correlation objective is to find the homologous point in the right image for point M. The first step in this process is to shift the same size region onto the right image until the correlation weight achieves the global maximum.
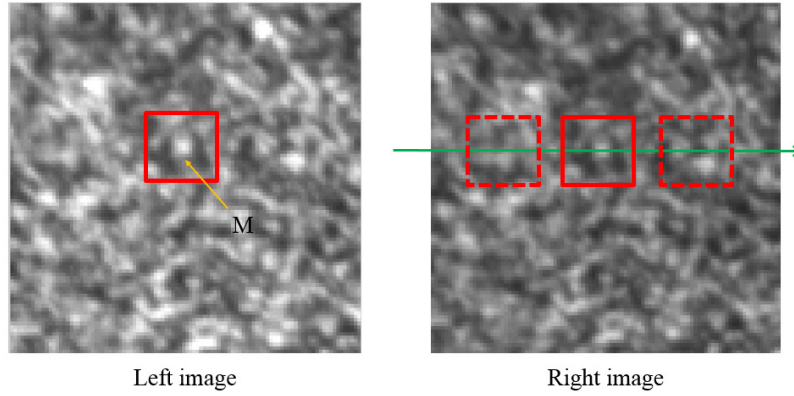

Left image   Right image

Fig. 2. Searching for a corresponding point

To evaluate the correlation degree between the left and right regions, a correlation criterion has to be established beforehand. At present, although correlation coefficient functions have many definitions and deformations, Pearson's product-moment correlation is more robust and easier to handle [4]. For two one-dimensional signals, f_L and f_R, the correlation coefficient ω is calculated using Equation (2):

$$\omega = \frac{cov(f_L, f_R)}{\sigma_L \sigma_R} = \frac{\sum_{k=1}^{NS} (f_{Lk} - \overline{f_{LS}})(f_{Rk} - \overline{f_{RS}})}{\sqrt{\sum_{k=1}^{NS} (f_{Lk} - \overline{f_{LS}})^2} \sqrt{\sum_{k=1}^{NS} (f_{Rk} - \overline{f_{RS}})^2}} \tag{2}$$

where NS is the size of the region f_{LS} and f_{RS} in the operation, and $\overline{f}$ is the mean value [3-4].

It should be noted that as it takes a lot of time to search the global maximum, many methods are proposed to increase the searching time, in this paper, we have applied methods of projection correction, parallax constraint and gray constraint to increase the speed [5].

2.3 Subpixel search method

Integer pixel correspondence was established in the previous section. To acquire high-precision measurement results, it is necessary to conduct subpixel interpolation. According to the spatial correlation method, we employ the second order disparity model [6] to describe the pixel coordinate bias in the spatiotemporal method, as shown in Equation (3):

$$\begin{cases} u^R = u^L + u_d + \frac{\partial u_d}{\partial u} \Delta u + \frac{\partial u_d}{\partial v} \Delta v + \frac{1}{2} \frac{\partial^2 u_d}{\partial u^2} (\Delta u)^2 + \frac{1}{2} \frac{\partial^2 u_d}{\partial v^2} (\Delta v)^2 + \frac{1}{2} \frac{\partial^2 u_d}{\partial u \partial v} \Delta u \Delta v \\ v^R = v^L + v_d + \frac{\partial v_d}{\partial u} \Delta u + \frac{\partial v_d}{\partial v} \Delta v + \frac{1}{2} \frac{\partial^2 v_d}{\partial u^2} (\Delta u)^2 + \frac{1}{2} \frac{\partial^2 v_d}{\partial v^2} (\Delta v)^2 + \frac{1}{2} \frac{\partial^2 v_d}{\partial u \partial v} \Delta u \Delta v \end{cases} \tag{3}$$

where (u^R, v^R) is the searched point in the right camera corresponding to (u^L, v^L) in the left camera; Δu and Δv denote the shift between the point (u^L, v^L) and the other pixels in the left image region; u_d and v_d are the zero order disparities; $\partial u_d/\partial u, \partial u_d/\partial v, \partial v_d/\partial u$, and $\partial v_d/\partial v$ denote the gray gradients in the chosen region; and $\partial^2 u_d/\partial u^2$, $\partial^2 u_d/\partial v^2$, $\partial^2 u_d/\partial u \partial v, \partial^2 v_d/\partial u^2, \partial^2 v_d/\partial v^2$, and $\partial^2 v_d/\partial u \partial v$ denote the second derivatives of the region. Next, a parameter vector $\boldsymbol{s}$, which contains all these twelve coefficients, is defined:

$$\boldsymbol{s} = (u_d, v_d, \frac{\partial u_d}{\partial u}, \frac{\partial u_d}{\partial v}, \frac{\partial v_d}{\partial u}, \frac{\partial v_d}{\partial v}, \frac{\partial^2 u_d}{\partial u^2}, \frac{\partial^2 u_d}{\partial v^2}, \frac{\partial^2 u_d}{\partial u \partial v}, \frac{\partial^2 v_d}{\partial u^2}, \frac{\partial^2 v_d}{\partial v^2}, \frac{\partial^2 v_d}{\partial u \partial v}) \tag{4}$$

The subpixel interpolation is to find a proper (u^R, v^R) that makes the correlation coefficient with (u^L, v^L) reach a maximum. According to Equations (3) and (4), the problem is translated into a search for a proper $\boldsymbol{s}$, which is a nonlinear optimization problem. In digital image correlation methods, Newton-Raphson iteration is a partial differential correlation approach, which has been proven to have high subpixel accuracy [7-9]. Combining Equations (2)–(4), the Newton-Raphson iterative solution at the k-th interval can be written as follows:

$$\boldsymbol{s}_k = \boldsymbol{s}_{k-1} - \frac{\nabla \omega(s)}{\nabla \nabla \omega(s)} \tag{5}$$

where

$$\nabla(\omega) = \sum_{u=-w_m}^{w_m} \sum_{v=-w_m}^{w_m} \left\{ \left[\frac{I_t^L(u^L, v^L) - \overline{I_t^L}}{\sqrt{\sum_{u=-w_m}^{w_m} \sum_{v=-w_m}^{w_m} [I_t^L(u^L, v^L) - \overline{I_t^L}]^2}} - \frac{I_t^R(u^R, v^R) - \overline{I_t^R}}{\sqrt{\sum_{u=-w_m}^{w_m} \sum_{v=-w_m}^{w_m} [I_t^R(u^R, v^R) - \overline{I_t^R}]^2}} \right] \cdot \right.$$

$$\left. \frac{1}{\sqrt{\sum_{u=-w_m}^{w_m} \sum_{v=-w_m}^{w_m} [I_t^L(u^L, v^L) - \overline{I_t^L}]^2}} \cdot \frac{\partial I_t^R(u^R, v^R)}{\partial s_i} \right\}_{i=1,2,\ldots,ns} ,$$

$$\nabla \nabla(\omega) = \frac{1}{\sum_{u=-w_m}^{w_m} \sum_{v=-w_m}^{w_m} [I_t^R(u^R, v^R) - \overline{I_t^R}]^2} \cdot \sum_{u=-w_m}^{w_m} \sum_{v=-w_m}^{w_m} \left\{ \frac{\partial I_t^R(u^R, v^R)}{\partial s_i} \cdot \frac{\partial I_t^R(u^R, v^R)}{\partial s_j} \right\}_{\substack{i=1,2,\ldots,ns \\ j=1,2,\ldots,ns}} ,$$

2.4 Registration of multi node 3D data

In the multi-node measurement network, each node obtains three dimensional graphs from different perspectives, to obtain the whole 3D foot data, range data registration is another key issue. Before capturing the depth data, the measurement network is calibrated, which means the transformation matrixes are obtained beforehand.

Take two viewpoints as an example. Assume data of the two viewpoints are $\mathbf{P} = \{\mathbf{p_1}, \mathbf{p_2}, \cdots, \mathbf{p_n}\}$ and $\mathbf{Q} = \{\mathbf{q_1}, \mathbf{q_2}, \cdots, \mathbf{q_n}\}$. The essence of registration is to transform the 3D depth data measured from different coordinate systems into a unified coordinate system. The transformation matrix A can be decomposed into a rotation matrix R and a translation matrix T. Influenced by measurement errors, M can be solved, which is equivalent to an optimization problem [10].

$$\boldsymbol{min} = \sum_{i=1}^{n} \|\mathbf{p}_i - \boldsymbol{A} \cdot \mathbf{q}_i\|^2 \tag{6}$$

where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}$, it can be solved by SVD method.

For multiple node measurement system, assume that $\boldsymbol{A_j}$ is the transformation matrix of j-th viewpoint data $\boldsymbol{V_i}$ from viewpoint coordinate frame to the world coordinate frame. The optimization problem can be denoted by:

$$\boldsymbol{min} = \sum_{j=1}^{K} \sum_{i=1}^{n} \|\mathbf{V}_{wi} - \boldsymbol{A}_k \cdot \mathbf{V}_i\|^2 \tag{7}$$

where K is node number, $\mathbf{V}_{wi}$ is the i-th coordinate in the world coordinate frame. The convergence condition can be converted to that there is a $\Delta\boldsymbol{A_k}$, which satisfies:

$$\begin{cases} \|\Delta\boldsymbol{A_k} - \boldsymbol{I}\| < \varepsilon \\ \boldsymbol{A}_{k+1} = \Delta\boldsymbol{A_k}\boldsymbol{A_k} \end{cases} \tag{8}$$

where $\boldsymbol{I}$ is the unit matrix, ε is the set threshold.

3. Experiment results

A multiple node 3D foot measurement system is constructed as shown in figure 3, including three nodes. The system contains six cameras and two digital projectors. After calibration of the cameras and the network, random speckles are projected on the foot model, within 1 second, the data is acquired by six cameras. Based on the stereo vision principle, 3D data of each viewpoint is acquired. With calibration parameters of the network, we can finally obtain the whole 3D model.

Fig. 3: System apparatus and software

In the model, the density of the point cloud is larger, and the data contains more redundant points, which is not conducive to data storage and display. The voxel grid and down sampling method can reduce the number of 3D points while maintaining the shape characteristics of point clouds. The three dimensional point cloud is subdivided into several small cubes, and the center of gravity of each small cube is calculated. All the points in the small cube are approximated by the gravity center, and the three dimensional point cloud data is obtained after the down sampling. Also, the number of vertices is related to the size of voxel grid.
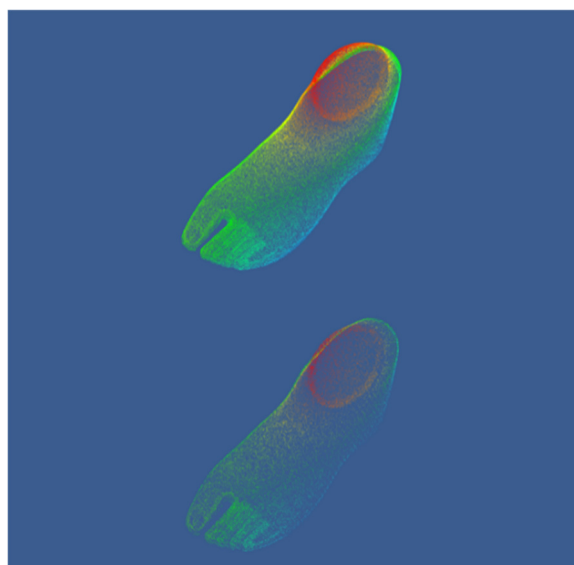

Fig. 4. Voxel grid and down sampling

The 3D model before and after the voxel grid is shown in figure 4. It can be seen that the down sampling method under the voxel grid can reduce the number of three dimensional points while maintaining the shape characteristics of the point cloud. It can improve the computing speed in shape recognition, surface reconstruction and matching algorithm. And the final reconstruction results are presented in figure 5.
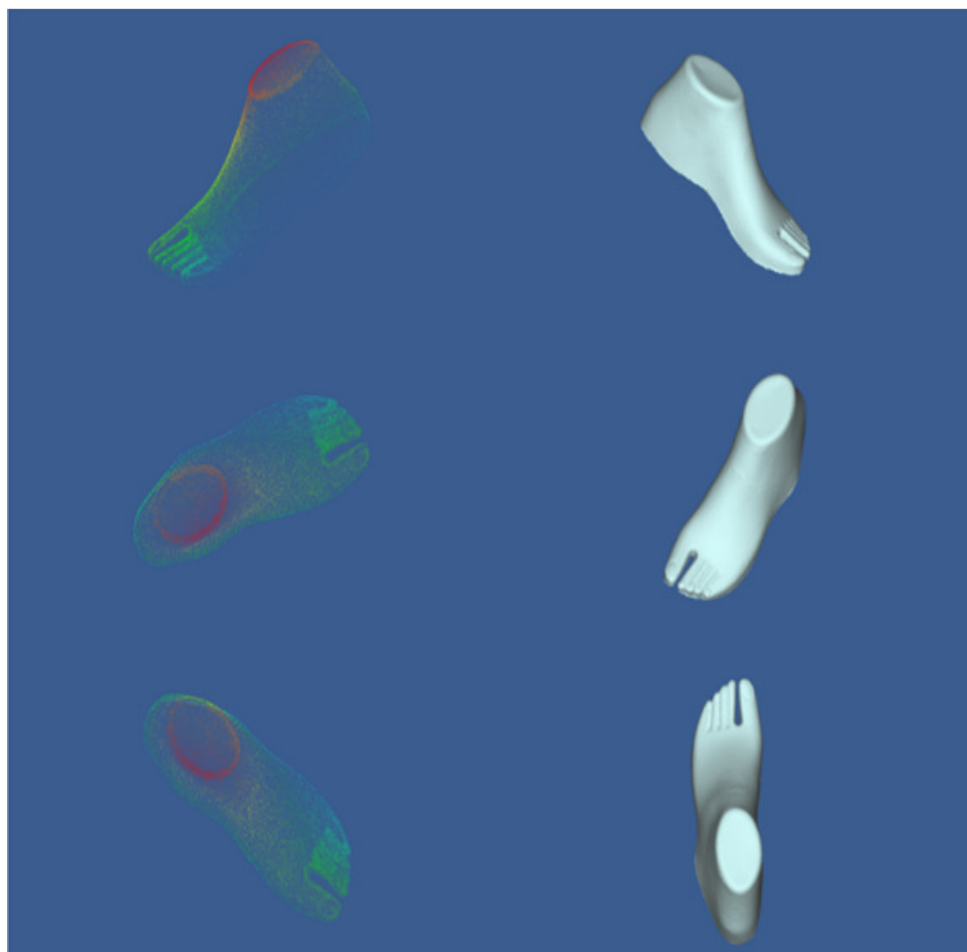
Fig. 5. Reconstruction result

4. Conclusions

The 3D reconstruction based on digital speckle correlation method has the advantage of single pattern reconstruction. In this paper, a multi-node 3D measurement system is applied for 3D foot reconstruction. On the basis of projection correction, using parallax constraint and gray constraint, the calculation times of each measured point are reduced, which greatly improves the search efficiency of the corresponding point. The matching and fusion of multi-view depth data is achieved by using measuring network calibration technology. The data acquisition speed of 3D foot measurement system has reached less than one second. The output of the 3D foot point cloud data can further realize the measurement of the key sizes of the human foot, and provide effective data support for the application of foot analysis and personalized shoe making.

Acknowledgement

References

[1]   J. Guo, et. al, "Automatic and rapid whole-body 3D shape measurement based on multinode 3D sensing and speckle projection", Appl Opt, Vol.56, No.31, 2017, pp. 8759-8768.

[2]   G. Sansoni, et. al, "State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation", Sensors, Vol.9, No.1, 2009, pp. 568-601.

[3]   Q. Tang, et. al, "An improved spatiotemporal correlation method for high- accuracy random speckle 3D reconstruction", Optics and Lasers in Engineering, Vol. 110, 2018, pp.54-62.

[4]   B. Pan, et. al, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: A review", Meas Sci Technol, Vol.20, 2009, pp. 062001.

[5]   Q. Tang, et. al, " Analysis of speckle characteristics in three-dimensional reconstruction based on digital speckle", Chinese Journal of Lasers, Vol.48, No.10, 2018, pp.1004004.

[6]   H. Schreier, et. al, "Systematic errors in digital image correlation due to undermatched subset shape functions", Exp Mech, Vol.42, No.3, 2002, pp. 303-10.

[7]   H. Bruck, et. al, "Digital image correlation using Newton-Raphson method of partial differential correction", Exp Mech, Vol.29, 1989, pp. 261–267.

[8]   C Cofaru, et. al, "Improved Newton-Raphson digital image correlation method for full-field displacement and strain calculation", Appl Opt, Vol.49, 2010, pp. 6472–84.

[9]   C Cofaru, et. al, "Pixel-level robust digital image correlation", Opt Express, Vol.21, No.24, 2013, pp.   29979-99.

[10]   X. Liu, et. al, "A method for global registration of range data combined with markers", Acta Optica Sinica, Vol.29, No.1, 2009, pp.1010-1014.