J. Meas. Sci. Instrum., 2024, 15(2): 166-175.
DOI: 10.62756/jmsi.1674-8042.2024017
http://jmsi.publish.founderss.cn
jmsi@nuc.edu.cn

A method for obtaining complete point cloud of tongue surface based on fringe projection trinocular system

WANG Peng*, ZHANG Xue, SUN Changku, FU Luhua

State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China

***Corresponding author:** WANG Peng (wang_peng@tju.edu.cn)

Received: January 7, 2024 **Revised:** March 5, 2024 **Accepted:** April 12, 2024

Abstract: The traditional tongue diagnosis process has the problem of poor objectivity. Applying computer vision technology to tongue diagnosis can effectively promote its objectivity. Binocular stereo vision combined with structured light fringe projection technology is a common method for 3D measurement. However, in the measurement scenario of tongue diagnosis, due to the presence of saliva and fluids on the tongue surface, there are high-reflectance areas with significant random distribution in the fringe images, leading to errors in phase calculation and point cloud loss. A trinocular measurement system was proposed based on fringe projection, where a trinocular system and three binocular subsystems were composed of three cameras. Dual-epipolar constraint based on phase and order constraints was introduced to enhance the accuracy of trinocular stereo matching. Supplementary matching points were utilized to optimize the trinocular matching point sets, reconstructing point clouds in high-reflectance areas. The results indicated that, compared to traditional binocular systems, this system achieved improved matching and reconstruction accuracy. Particularly in real tongue surface measurements, it could generate point clouds with clear textures and complete features. It could effectively measure the highly reflective area of the tongue surface and facilitate objective tongue diagnosis.

Key words: 3D measurement; trinocular system; fringe projection; dual-epipolar constraint

0 Introduction

Digitizing and visualizing the 3D morphology of the tongue surface, and accurately extracting and measuring medical features, are of significant importance for advancing the objectification of tongue diagnosis. Grating fringe projection has advantages such as high accuracy and fast speed[1]. However, in 3D measurement, it requires phase unwrapping of the fringe images. This process is more easily accomplished when the measured object has good diffuse reflection characteristics. Due to the application environment of 3D tongue diagnosis, the tongue surface to be measured is often accompanied by saliva or fluids, leading to the presence of high-reflectance areas or brightness saturation areas with significant random distribution in the images captured by the camera. It results in errors in phase calculation and the loss of 3D reconstructed point clouds.

For the 3D measurement of highly reflective surfaces, there have been significant research advancements. Firstly, the multiple-exposure method has been explored[2,3]. Jiang et al.[4] analyzed the characteristics of reflected light on highly reflective surfaces and its impact on phase calculation. They proposed techniques such as bright-dark fringe projection and image acquisition with multiple exposure times, effectively addressing the issue of missing point clouds in measurements. This method requires capturing multiple sets of images, resulting in a longer measurement time, and strict requirements on the brightness of the fringes. Another approach is to vary the brightness of the projected fringes[5,6]. Liu et al.[7] proposed a method for active compensation of structured light saturation in a regionally projected manner. The method locates the saturated region by using the grayscale map of the Gray code in the projection region. Subsequently, it increases the transitional compensation region to smoothly reduce the intensity of the fringe projection, thereby suppressing reconstruction errors caused by brightness saturation. Lin et al.[8] introduced an adaptive digital fringe projection technique that adjusted the intensity of the projected pattern pixel by pixel, reducing image saturation. This method involves complex algorithms, has a large computational load, and consumes a significant amount of time. Another approach involves introducing auxiliary devices. For example, Zhao et al.[9] introduced a DMD camera to achieve adaptive 3D morphology measurement of HDR surface. The DMD modulation function is

determined before obtaining the phase-shift images, thus avoiding pixel saturation. Salahieh et al.[10] proposed a multi-polarized fringe projection (MPFP) imaging technique. This technique eliminates saturation and enhances fringe contrast by selecting appropriate polarization channels or the correct combination of polarization angles and exposure times, enabling more accurate morphological measurements. Riviere et al.[11] incorporated three different orientations of rotating linear polarizing filters in front of the camera to limit the polarization angle of reflected light from highly reflective surfaces, thereby reducing the likelihood of overexposure. This method increases the complexity of the system and reduces the intensity of the captured images. Lastly, another approach is the multi-view method[12]. Liu et al.[13] treated the monocular and projector in a binocular-projector system as a binocular structure. Multiple binocular systems simultaneously captured fringe images from different perspectives, and image segmentation techniques were applied to determine the optimal reconstruction area.

The surface reflection model[14] indicates that surface-reflected light includes diffuse reflection lobes, specular reflection lobes, and specular reflection peaks[15], and pixel saturation in images is due to the high intensity specular reflection peaks. Considering that specular reflection regions follow the reflection theorem, strong reflection points only experience exposure from a single viewpoint. Therefore, this paper adopted a trinocular vision system, utilizing the multi-view method for tongue surface acquisition. In the research on trinocular system, Ma et al.[16] designed a trinocular model with variable baseline and employed three binocular structures to cooperatively measure object dimensions. However, binocular matching may miss pixels in high-reflectance areas, resulting in incomplete reconstruction. Moreover, this method requires repeated calibration for different measurement objects, making the operation cumbersome.

Conen et al.[17] built an endoscopic system using three small camera modules, reconstructing features by aggregating cost volumes through three-camera matching. This method involves significant computational costs and is not effective for measuring highly reflective surfaces such as fluids. Xiao et al.[18] utilized a CCD camera and a pyramid-shaped prism to generate three virtual cameras, forming a trinocular system. While this system ensures synchronous capture, it also reduces the resolution of each image, resulting in a lower point cloud density during reconstruction.

Based on fringe projection for 3D measurement, a trinocular-projection measurement system was established, comprising a trinocular system and three binocular subsystems. The trinocular system captured images from different perspectives and completed the reconstruction of point clouds in the highly reflective areas of the tongue surface through trinocular stereo matching and supplementary matching points. Finally, a complete point cloud of the tongue surface was obtained through the fusion of point clouds from multiple systems.

1 Measurement system model

The trinocular measurement system model proposed in this paper is shown in Fig.1. Taking the left camera as an example, its camera coordinate system is $O_{cl}-X_{cl}Y_{cl}Z_{cl}$, and the pixel coordinate system is $o_l-u_lv_l$. A point $\boldsymbol{P}_w(x_w,y_w,z_w)$ in the world coordinate system $O_w-X_wY_wZ_w$ corresponds to points $\boldsymbol{p}_l(u_l,v_l)$, $\boldsymbol{p}_m(u_m,v_m)$ and $\boldsymbol{p}_r(u_r,v_r)$ in the pixel coordinate systems of the left, middle, and right cameras, respectively.

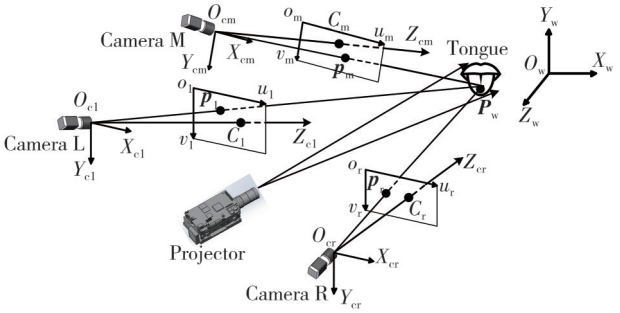


Fig. 1 Schematic diagram of trinocular measurement system

According to the principles of stereo vision[19], in the trinocular system, the transformation relationship from the world coordinate system to the pixel coordinate system is

$$\begin{cases} Z_{cl}\boldsymbol{p}_l=\boldsymbol{M}_l\boldsymbol{P}_w, \\ Z_{cm}\boldsymbol{p}_m=\boldsymbol{M}_m\boldsymbol{P}_w, \\ Z_{cr}\boldsymbol{p}_r=\boldsymbol{M}_r\boldsymbol{P}_w, \end{cases} \tag{1}$$

where Z_{cl}, Z_{cm}, and Z_{cr} are the Z_c axis coordinates of point $\boldsymbol{P}_w$, and normalization is performed during calculation. $\boldsymbol{M}_l$, $\boldsymbol{M}_m$, and $\boldsymbol{M}_r$ are the projection matrices of the left, middle, and right cameras, which can be obtained through camera calibration. By using trinocular stereo matching, the pixel coordinates of three matching points can be obtained, and then the coordinates of point $\boldsymbol{P}_w$ can be calculated.

In addition, the three cameras were paired to form the left-middle, right-middle, and left-right binocular subsystems (abbreviated as systems LM, MR, and LR), respectively. These three subsystems reconstructed the tongue surface from different perspectives, compensating for the lack of edge information in a single view, thus

supplementing the reconstructed point cloud of the trinocular system and ensuring the completeness of tongue surface reconstruction. To avoid structural misalignment in the resulting point cloud, it is necessary to unify the coordinate systems of the multiple systems.

As shown in Fig.2, for ease of calculation, the world coordinate systems of systems LM, LR, and the trinocular system are unified to the left camera coordinate system. $\boldsymbol{R}_{\mathrm{l}}$ and $\boldsymbol{T}_{\mathrm{l}}$ are defined respectively as the rotation and translation matrices from the world coordinate system to the left camera coordinate system, where $\boldsymbol{R}_{\mathrm{l}}=\boldsymbol{E}$ and $\boldsymbol{T}_{\mathrm{l}}=\boldsymbol{0}$.

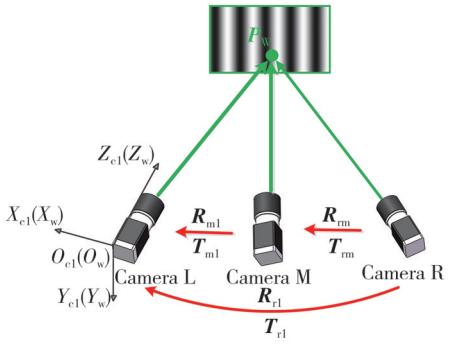


Fig. 2 Coordinate system transformation for multiple systems

With the middle camera as the reference in system MR, the world coordinate point $\boldsymbol{P}_{\mathrm{mw}}$ in this system can be transformed into the coordinate system based on the left camera using Eq. (2), resulting in point $\boldsymbol{P}'_{\mathrm{mw}}$, thus achieving coordinate system unification across multiple systems, that is

$$\boldsymbol{P}'_{\mathrm{mw}}=\boldsymbol{R}_{\mathrm{ml}}\boldsymbol{P}_{\mathrm{mw}}+\boldsymbol{T}_{\mathrm{ml}}, \tag{2}$$

where $\boldsymbol{R}_{\mathrm{ml}}$ and $\boldsymbol{T}_{\mathrm{ml}}$ represent the rigid transformation relationship between the middle and left camera coordinate systems, obtained through stereo calibration experiments.

2 Trinocular stereo matching and reconstruction in high-reflectance areas

2.1 Extraction of fringe phase features

The order and phase information contained in the fringe images are eigenvalues for stereo matching. By projecting three sets of sinusoidal fringes with different frequencies (each set including four fringe patterns with phase variations at equal intervals), 12 fringe feature images were obtained. The main phase of each group of fringe images, namely the wrapping phase $\varphi(x,y)$, is obtained by applying the phase calculation algorithm, as shown below.

$$\varphi(x,y)=\arctan\left[\frac{\sum_{k=1}^{N}I_k(x,y)\sin(2k\pi/N)}{\sum_{k=1}^{N}I_k(x,y)\cos(2k\pi/N)}\right], \tag{3}$$

where $N=4$, $I_k(x,y)$ is the pixel intensity value at coordinates (x,y) in the kth image. The optimum frequency selection method[20] was introduced for phase unwrapping. Three sets of fringe frequencies are defined as f_0, f_1, and f_2, and the difference frequency operation is defined as

$$\begin{cases} f_{02}=f_0-f_2, \\ f_{12}=f_1-f_2, \\ f^*=f_{02}-f_{12}. \end{cases} \tag{4}$$

The wavelength of the difference frequency f^* is $\lambda^*=\lambda_{02}\lambda_{12}/(\lambda_{12}-\lambda_{02})$, where λ_{02} and λ_{12} are the wavelengths corresponding to f_{02} and f_{12}. Since λ^* is the wavelength of a complete cycle and is known, let the distance from a certain phase point to the zero phase point be d, then

$$d=(m^*+\varphi^*/2\pi)\lambda^*=(m_0+\varphi_0/2\pi)\lambda_0, \tag{5}$$

where m^* and φ^* are the order and principal phase of fringes at frequency f^*, m_0 and φ_0 are the order and principal phase of fringes at frequency f_0. Since $\lambda^*>d$, it follows that $m^*=0$. Thus, the order m_0 of this phase point is determined as

$$m_0=\mathrm{INT}\left[\frac{\varphi^*}{2\pi}\frac{\lambda^*}{\lambda_0}-\frac{\varphi_0}{2\pi}\right]. \tag{6}$$

Therefore, the continuously distributed unwrapped phase in pixel coordinate system is

$$\phi(x,y)=\varphi_0(x,y)+2\pi m_0(x,y). \tag{7}$$

2.2 Stereo matching with dual-epipolar constraint

After obtaining continuous phase and order information, it is necessary to introduce epipolar constraint for trinocular stereo matching. According to the epipolar geometry principle[21], every point in one image obtained by two cameras simultaneously capturing an object corresponds to a line in another image, known as the epipolar line for that point. Fig.3 shows the epipolar geometry principle in the trinocular system.

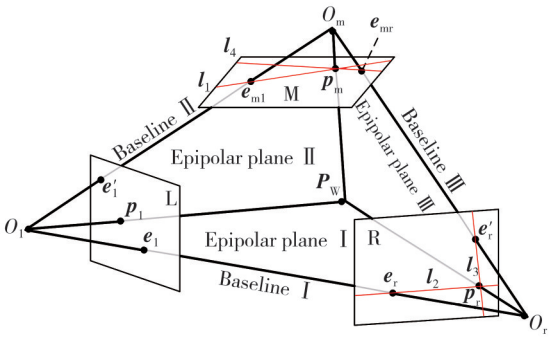


Fig. 3 Epipolar geometry principle in trinocular system

As shown in Fig.3, in a trinocular system, when three cameras simultaneously capture an object point $\boldsymbol{P}_{\mathrm{w}}$, the

right matching point $\boldsymbol{p}_r$ must lie on the right epipolar line $\boldsymbol{l}_2$ corresponding to the left matching point $\boldsymbol{p}_l$. The epipolar constraint equation can be expressed as

$$\boldsymbol{p}_r^{\mathrm{T}}\left(\boldsymbol{I}_r^{-1}\right)^{\mathrm{T}}\boldsymbol{E}_{12}\boldsymbol{I}_l^{-1}\boldsymbol{p}_l=\boldsymbol{p}_r^{\mathrm{T}}\boldsymbol{F}_{12}\boldsymbol{p}_l=0, \tag{8}$$

where $\boldsymbol{I}_l$ and $\boldsymbol{I}_r$ are the intrinsic matrices of the left and right cameras; $\boldsymbol{E}_{12}$ is the essential matrix between the left and right cameras, and $\boldsymbol{F}_{12}=\left(\boldsymbol{I}_r^{-1}\right)^{\mathrm{T}}\boldsymbol{E}_{12}\boldsymbol{I}_l^{-1}$ is the fundamental matrix between the left and right cameras. Therefore, the left and right epipolar lines $\boldsymbol{L}_l$ and $\boldsymbol{L}_r$ can be represented as

$$\begin{cases}\boldsymbol{L}_l=\boldsymbol{I}_l\boldsymbol{F}_{12}\boldsymbol{F}_{12}^{\mathrm{T}}\boldsymbol{p}_r,\\ \boldsymbol{L}_r=\boldsymbol{I}_r\boldsymbol{F}_{12}^{\mathrm{T}}\boldsymbol{F}_{12}\boldsymbol{p}_l,\end{cases} \tag{9}$$

where $\boldsymbol{L}_r$ represents the epipolar line $\boldsymbol{l}_2$ on the right image plane $\boldsymbol{R}$; similarly, the matching point $\boldsymbol{p}_m$ on the middle image plane $\boldsymbol{M}$ lies on the epipolar line $\boldsymbol{l}_1$.

Traditional binocular systems can only constrain stereo matching through a single epipolar line, and the single epipolar constraint may lead to situations where entire pixel points cannot be completely matched. Linear interpolation operations are needed to calculate sub-pixel coordinates, indirectly affecting the accuracy of pixel stereo matching. Leveraging the advantages of the trinocular system, a second epipolar line through the third camera was introduced, forming a dual-epipolar constraint to eliminate matching ambiguity and enhance matching accuracy. The matching process is illustrated in Fig.4.

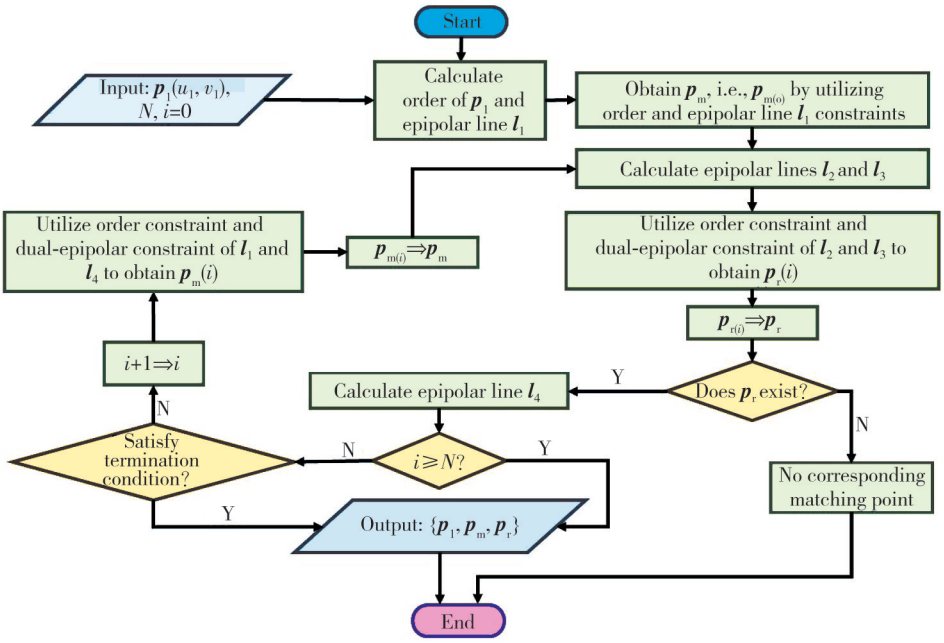


Fig. 4 Flowchart of stereo matching with dual-epipolar constraints

First, a point on the left image plane was calculated as a epipolar line on the middle image plane according to the epipolar formula.

$$\boldsymbol{L}_m=\boldsymbol{I}_m\boldsymbol{F}_{10}^{\mathrm{T}}\boldsymbol{F}_{10}\boldsymbol{p}_l, \tag{10}$$

where $\boldsymbol{I}_m$ is the intrinsic matrix of the middle camera; $\boldsymbol{F}_{10}$ is the fundamental matrix between the left and middle cameras; and $\boldsymbol{L}_m$ represents the epipolar line $\boldsymbol{l}_1$.

Then, in the left and middle stereo images, the matching point $\boldsymbol{p}_m$ for the point $\boldsymbol{p}_l$ is found using order constraint and the epipolar constrainxct of $\boldsymbol{l}_1$. Similarly, for the point $\boldsymbol{p}_m$, introducing the fundamental matrix $\boldsymbol{F}_{02}$ between the right and middle cameras can obtain its epipolar line $\boldsymbol{l}_3$ on the right image plane $\boldsymbol{R}$. Combining the epipolar line $\boldsymbol{l}_2$ calculated in Eq. (9), $\boldsymbol{l}_2$ and $\boldsymbol{l}_3$ form a dual-epipolar constraint for matching point $\boldsymbol{p}_r(u_r, v_r)$ on the right image corresponding to point $\boldsymbol{p}_l$. Combining order constraint, the constraint point set $\boldsymbol{P}_r$ for the right matching point $\boldsymbol{p}_r$ can be represented as

$$\begin{cases}\left\{\boldsymbol{P}_r=(\boldsymbol{P}_1\cap\boldsymbol{P}_2)\cap\boldsymbol{P}_3\right\},\\ \left\{\boldsymbol{P}_1|(u_r,v_r)\in\boldsymbol{P}_1,\left|a_2u_r+b_2v_r+c_2\right|<\delta\right\},\\ \left\{\boldsymbol{P}_2|(u_r,v_r)\in\boldsymbol{P}_2,\left|a_3u_r+b_3v_r+c_3\right|<\delta\right\},\\ \left\{\boldsymbol{P}_3|m_r(\boldsymbol{P}_1)=m_r(\boldsymbol{P}_2)=m_l[(u_l,v_l)]\right\},\end{cases} \tag{11}$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are the point sets on the corresponding epipolar lines $\boldsymbol{l}_2$ and $\boldsymbol{l}_3$; a_2, b_2, c_2 and a_3, b_3, c_3 are the parameters of the epipolar lines $\boldsymbol{l}_2$ and $\boldsymbol{l}_3$, respectively; δ is the set minimum threshold. Thus, the initial trinocular matching point set $\{\boldsymbol{p}_l, \boldsymbol{p}_m, \boldsymbol{p}_r\}$ is obtained.

To optimize the matching results, an iterative approach was adopted through multiple rounds of stereo matching with dual-epipolar constraint to refine the initial matching results and gradually improve matching accuracy. Initially, the formula $\boldsymbol{L}_m'=\boldsymbol{I}_m\boldsymbol{F}_{02}\boldsymbol{F}_{02}^{\mathrm{T}}\boldsymbol{p}_r$ was

used to calculate the epipolar line $\boldsymbol{l}_4$ on the middle image plane for the initial right matching point $\boldsymbol{p}_r$. The epipolar lines $\boldsymbol{l}_4$ and $\boldsymbol{l}_1$ together form dual-epipolar constraint for the matching point in the middle camera image, leading to the update of the matching point $\boldsymbol{p}_m$, and then enter the iteration process. Considering computational efficiency and the issue of over-fitting, the iteration is set with a fixed number N and additional termination conditions.

$$\begin{cases} i \geqslant 1, \\ \left| \boldsymbol{p}_{m(i)} - \boldsymbol{p}_{m(i-1)} \right| < \varepsilon, \\ \left| \boldsymbol{p}_{r(i)} - \boldsymbol{p}_{r(i-1)} \right| < \varepsilon, \end{cases} \tag{12}$$

where $\boldsymbol{p}_{m(i)}$ and $\boldsymbol{p}_{r(i)}$ are the matching results after the ith iteration, $i = 0,1,2,\cdots,N$. Iteration terminates when the change in matching points is less than the set distance threshold ε, and the matching point set is outputted.

2.3 Supplementary matching point set in high-reflectance areas

In this paper, starting from the minimum order point in the fringe image, all pixel points were traversed, and trinocular images stereo matching was performed using dual-epipolar and order constraints. For a measurement object with a surface exhibiting good diffuse reflection characteristics, ideally all object points should find corresponding matching points on the image planes of the three cameras. However, for the surface of the tongue, the matching situation is illustrated in Fig.5 due to the influence of high-reflectance factors.

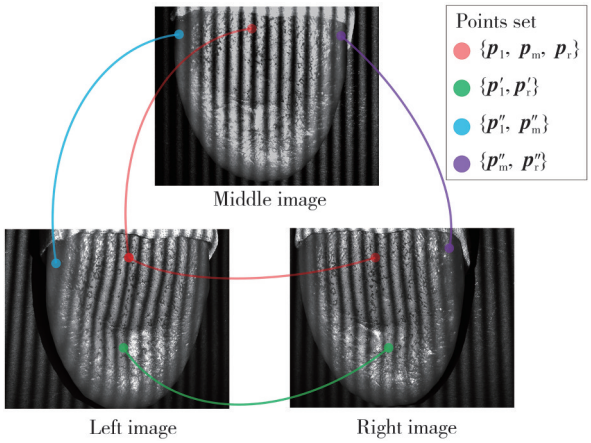


Fig. 5 Schematic diagram of feature point matching in trinocular system

In Fig. 5, the red set of points represents the set of matching points obtained under normal circumstances. In the other three cases, the relevant order and phase information can not be obtained due to the presence of highly saturated pixels in one of the camera images. Therefore, the matching points on the image plane of the camera cannot be located. When performing stereo matching on a plane containing overexposed regions, five specific cases may occur.

1) Matching points are obtained in the left, middle, and right camera images, and the matching point set is defined as $\boldsymbol{P}_{LMR}$:$\left\{\left\{ \boldsymbol{p}_l, \boldsymbol{p}_m, \boldsymbol{p}_r \right\}\right\}$.

2) Matching points are obtained only in the left and right camera images, and the matching point set is defined as $\boldsymbol{P}_{LR}$:$\left\{\left\{ \boldsymbol{p}'_l, \boldsymbol{p}'_r \right\}\right\}$.

3) Matching points are obtained only in the left and middle camera images, and the matching point set is defined as $\boldsymbol{P}_{LM}$:$\left\{\left\{ \boldsymbol{p}''_l, \boldsymbol{p}'_m \right\}\right\}$.

4) Matching points are obtained only in the right and middle camera images, and the matching point set is defined as $\boldsymbol{P}_{MR}$:$\left\{\left\{ \boldsymbol{p}''_m, \boldsymbol{p}''_r \right\}\right\}$.

5) Matching points obtained in only one camera image or no matching points are obtained in any image, meaning no information about the tongue surface is present in that point set. The epipolar constraint models for the first four cases are shown in Fig.6.

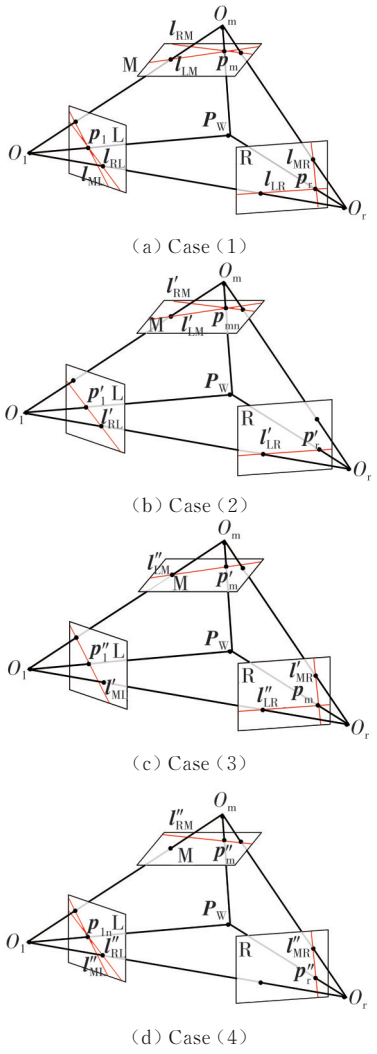


Fig. 6 Epipolar constraint models for four cases

Matching points could not be obtained in all three camera images, and the unknown matching point on the third image plane was supplemented by using the known epipolar line information of two matching points. Fig. 7 illustrates the situation in Fig. 6(c) in the phase unwrapping diagram.

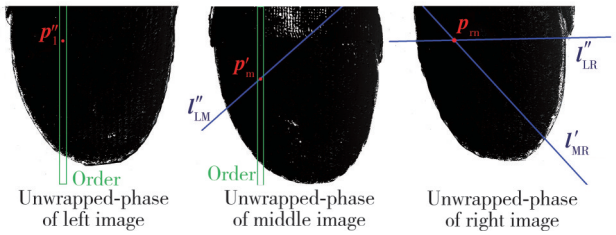


Fig. 7 Calculating supplementary matching point in phase unwrapping diagram

Utilizing the point sets $\{\boldsymbol{p}_{\mathrm{l}}''\}$ and $\{\boldsymbol{p}_{\mathrm{m}}'\}$ and their corresponding epipolar line sets $\{\boldsymbol{l}_{\mathrm{LR}}''\}$ and $\{\boldsymbol{l}_{\mathrm{MR}}'\}$ in the right image, the intersection of the two epipolar lines are calculated for each matching point pair using Eq. (13), resulting in the new right image matching point set $\{\boldsymbol{p}_{\mathrm{rn}}\}$.

$$\begin{cases} \boldsymbol{l}_{\mathrm{LR}}'' = \boldsymbol{I}_{\mathrm{r}} \boldsymbol{F}_{12}^{\mathrm{T}} \boldsymbol{F}_{12} \boldsymbol{p}_{\mathrm{l}}'', \\ \boldsymbol{l}_{\mathrm{MR}}' = \boldsymbol{I}_{\mathrm{r}} \boldsymbol{F}_{02}^{\mathrm{T}} \boldsymbol{F}_{02} \boldsymbol{p}_{\mathrm{m}}'. \end{cases} \tag{13}$$

The obtained supplementary matching point set for the trinocular images is defined as $\boldsymbol{P}_{\mathrm{LM}}'$:$\{\{\boldsymbol{p}_{\mathrm{l}}'', \boldsymbol{p}_{\mathrm{m}}', \boldsymbol{p}_{\mathrm{rn}}\}\}$, and it is added to $\boldsymbol{P}_{\mathrm{LMR}}$. Similarly, the solution is completed for the point sets $\{\boldsymbol{p}_{\mathrm{mn}}\}$ and $\{\boldsymbol{p}_{\mathrm{ln}}\}$ in Fig.6(b) and 6(d), achieving the supplementation of matching point sets in the high-reflectance areas of the tongue's trinocular images.

In reality, the element count of $\boldsymbol{P}_{\mathrm{LM}}'$ is reduced compared to $\boldsymbol{P}_{\mathrm{LM}}$, i.e., $|\boldsymbol{P}_{\mathrm{LM}}'| < |\boldsymbol{P}_{\mathrm{LM}}|$. This is due to the limitation of the camer's field of view, causing points on the edge of the tongue surface to be unable to obtain matching points in all three image planes. Therefore, this part of the points can only rely on the binocular subsystems for world coordinate points reconstruction, ultimately completing the supplementation of 3D information for the tongue surface through point cloud fusion.

2.4 Validation of supplementary matching point sets

To validate the matching quality of the trinocular images' supplementary matching point sets, the Census transform was employed for consistency verification[22]. The Census transform compares the relative grayscale relationships of pixels, tolerating certain levels of noise interference. Therefore, even if there is inconsistent brightness among the three camera images in practical measurements, the consistency verification of matching points can still be accomplished. Taking the matching point $\boldsymbol{p}_{\mathrm{ref}}$ as the center pixel, a rectangular window D is defined in the image area. Subsequently, the grayscale value of each pixel in the window is compared with the grayscale value of the center pixel, resulting in the Census transform result $T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{ref}})$ for the matching point. Taking the scenario in Fig. 6(c) as an example, as shown in Fig.8, the Census transform results for the left and middle matching points $\boldsymbol{p}_{\mathrm{l}}''$ and $\boldsymbol{p}_{\mathrm{m}}'$, as well as the supplementary right matching point $\boldsymbol{p}_{\mathrm{rn}}$, are denoted as $T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{l}})$, $T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{m}})$, and $T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{r}})$, respectively:

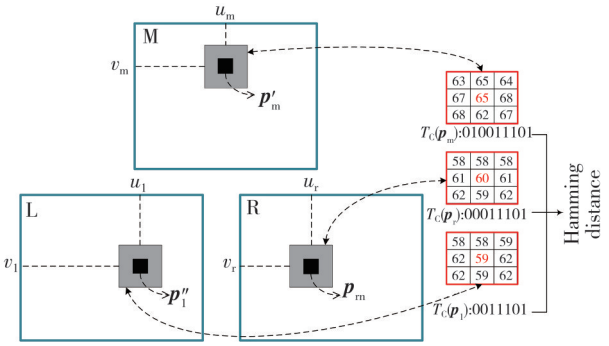


Fig. 8 Calculating Hamming distance between matching points

By calculating the Hamming distance d_{H} between matching points, the matching consistency was measured. A set of matching points passed the consistency verification under the condition.

$$\begin{cases} d_{\mathrm{H}}(T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{l}}), \quad T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{m}})) < d_{\mathrm{thr}}, \\ d_{\mathrm{H}}(T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{l}}), \quad T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{r}})) < d_{\mathrm{thr}}, \\ d_{\mathrm{H}}(T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{m}}), \quad T_{\mathrm{C}}(\boldsymbol{p}_{\mathrm{r}})) < d_{\mathrm{thr}}, \end{cases} \tag{14}$$

where d_{thr} is the Hamming distance threshold, supplementary matching points that do not satisfy the above conditions are considered as mismatches and are removed from the point set.

2.5 Point cloud generation

The matching point set $\boldsymbol{P}_{\mathrm{LMR}}$ was obtained by the trinocular stereo matching. By substituting the pixel coordinates of matching points into Eq. (1), the world coordinates $\boldsymbol{P}_{\mathrm{w}}(x_{\mathrm{w}}, y_{\mathrm{w}}, z_{\mathrm{w}})$ of the object points can be obtained. Color cameras were utilized to capture 2D color images for RGB three-channel mapping of the 3D point clouds, resulting in the initial color point cloud $C1$. As shown in Fig. 9, point cloud processing based on Poisson surface reconstruction[23] is applied to $C1$. This process fit a small amount of noise point cloud near the tongue surface onto the surface, resulting in the output point cloud $C5$ from the trinocular system.

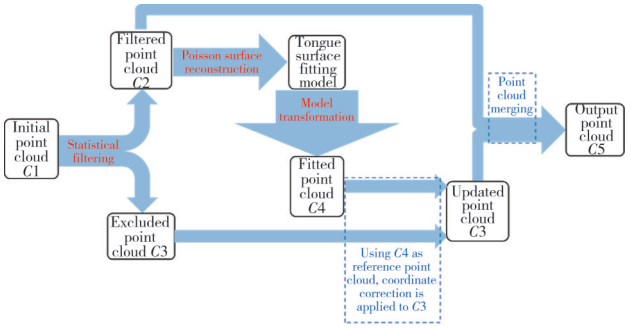


Fig. 9 Point cloud processing flowchart

3 Experiments

3.1 Experiment overview

The trinocular measurement system in this paper included three HIKVISION industrial cameras with a resolution of 1 280×1 024 pixels. The left and right cameras were MV-CA013-21UC color cameras, the middle camera was MV-CA013-21UM monochrome camera, and the projector was DLP4500 from TI. All three cameras used 16 mm lenses from HIKVISION. The baseline length was 150 mm for the left-right camera pair, 100 mm for the left-middle camera pair, and 100 mm for the right-middle camera pair. At a fixed working distance of 180 mm, the use of larger focal length lenses could concentrate the common field of view of the three cameras more on the tongue surface, thereby improving the point cloud density in the tongue area.

3.2 Stereo matching accuracy verification experiment

To verify the matching accuracy of the trinocular system with the introduction of dual-epipolar constraint, directional circular target was used as measurement object. The contour extraction algorithm was employed to calculate the pixel coordinates of the circle center points in the images, obtaining the true coordinate values of pixel stereo matching. Subsequently, the stereo matching results of both the trinocular system and the binocular system LR were obtained. This enabled a comparison of the matching accuracy between the two measurement systems. Figs.10 and 11 respectively show the matching situation of feature points in the left and right camera images of the trinocular system and the LR system, along with a locally enlarged schematic diagram. The line chart shown in Fig. 12 is obtained through several experiments.

Because the binocular matching accuracy was higher when the matching pixels were calculated as full pixels, there were fewer points where the matching error of the trinocular system was equal to or exceeded the matching error of the binocular system. The average matching error reduction ratio of the trinocular system was 10.3%, indicating that the matching method with dual-epipolar constraint could further enhance the accuracy of feature point matching.

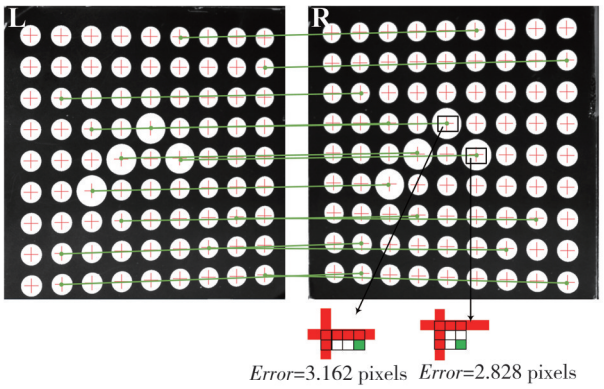


Fig. 10 Matching situation in left and right images of trinocular system

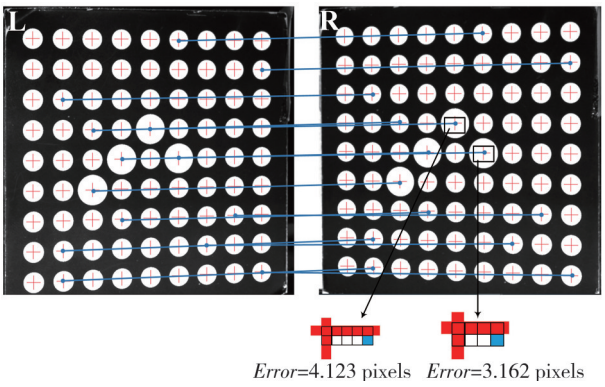


Fig. 11 Matching situation in images of LR system

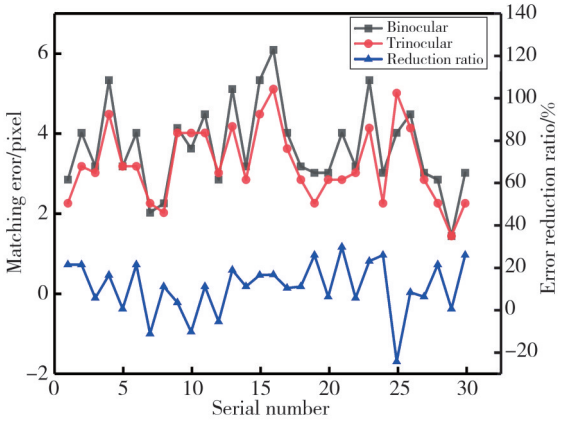


Fig. 12 Line chart of matching errors for trinocular and binocular systems and error reduction ratio

3.3 Accuracy verification experiment for 3D reconstruction

3.3.1 Verification of point cloud completion method

Before the reconstruction, it is necessary to supplement the matching points. Table 1 presents a

comparison of the supplementary matching points and the matching points validated through Census in multiple measurement experiments. The average pass ratio was 90.8%, indicating that the supplementary matching point sets exhibited good consistency and the effectiveness of the point cloud completion method proposed in this paper.

Table 1 Census verification status of supplementary matching points

Experiment number	Volume of supplementary matching points	Volume of points verified through Census	Pass ratio/%
1	39 261	35 989	91.7
2	40 990	37 768	92.1
3	42 503	37 937	89.3
4	38 723	36 047	93.1
5	45 190	40 971	90.7
6	51 096	46 054	90.1
7	40 112	37 099	92.5
8	39 482	34 251	86.8
Average pass ratio			90.8

3.3.2 Verification of reconstruction accuracy

On the basis of the effectiveness of the point cloud completion method, the reconstruction accuracy of the point cloud was verified through comparative experiments with traditional binocular reconstruction. In this experiment, the trinocular system and the binocular system were separately used to measure a standard sphere with a nominal diameter of 50.813 5 mm. The diameter was obtained through sphere fitting. Fig. 13 shows the diameter measurement results.

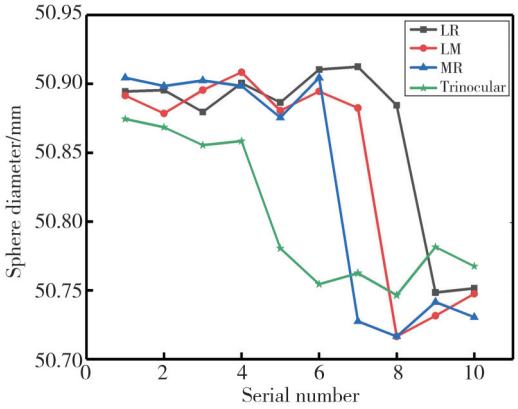


Fig. 13 Measurement results of standard sphere with different systems

Table 2 presents a comparison of various parameters, indicating that the reconstruction accuracy of the trinocular system is superior to that of the binocular system.

Table 2 Comparison of measurement results from different systems

System	Average of measurement results/mm	Maximum measurement deviation/mm	RMSE/mm
LR	50.866 4	0.098 3	0.079 3
LM	50.842 7	0.096 9	0.078 4
MR	50.863 5	0.090 5	0.084 6
Trinocular	50.806 9	0.067 7	0.050 3

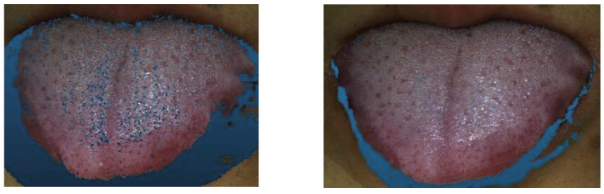
3.4 Measurement experiment of tongue surface

Multiple measurement experiments on the same tongue surface were conducted using both traditional binocular systems and the trinocular system proposed in this paper. Table 3 shows the data volume of the acquired point clouds and the growth ratio of point cloud data for the trinocular system compared to the binocular system. The average growth ratio was 6.55%, indicating that the method presented in this paper achieved denser point cloud reconstruction for tongue surface.

Table 3 Comparison of point cloud data volume between binocular and trinocular systems

Experiment number	Data volume of generated point cloud		Growth ratio/%
	Binocular system	Trinocular system	
1	727 728	781 080	7.33
2	799 839	855 267	6.93
3	780 169	831 036	6.52
4	856 523	908 856	6.11
5	756 547	808 067	6.81
6	675 799	720 063	6.54
7	798 472	845 691	5.91
8	733 916	779 712	6.24

Fig.14 shows the output point clouds of the traditional binocular system and the system proposed in this paper.



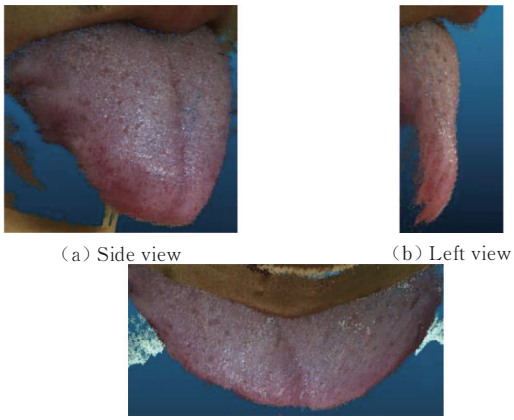
(a) Output of binocular system (b) Output of proposed system

Fig. 14 Comparison of point cloud reconstruction results

It could be observed that the output point cloud of the proposed system possessed characteristics such as clear tongue surface texture and complete features. The high-reflectance areas on the tongue surface were effectively supplemented. Fig. 15 are images of the tongue surface point cloud from other perspectives.

It could be observed that there were fewer scattered points near the tongue surface, indicating that the three-dimensional reconstruction effect of the proposed system on the tongue surface was relatively ideal. Fig.16 shows

the point cloud reconstruction results of various tongue models using the proposed system. It could be seen that various texture features were prominent, indicating the practical value of the system.



(a) Side view (b) Left view

(c) Top view

Fig. 15 Point clouds from various perspectives

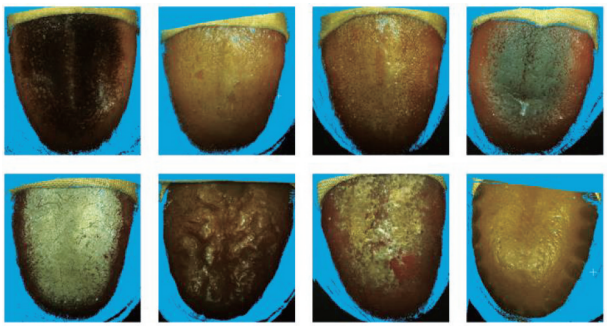
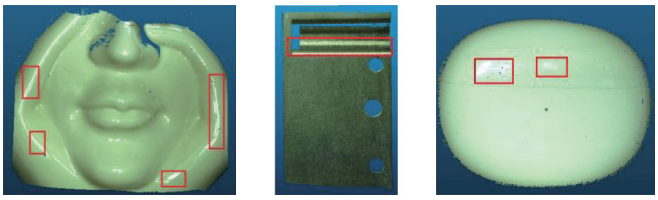


Fig. 16 Point cloud of various tongue models

Fig.17 shows the measurement results for other high-reflectance objects. It could be seen that the highly reflective feature reconstruction of the red marked area was complete, further verifying the effectiveness of the proposed method in the high-reflectance measurement.



(a) Ceramic artifact (b) Metal parts (c) Plastic shell

Fig. 17 Point cloud for different high-reflectance objects

Compared with the methods of multiple-exposure and variations in projected brightness, the proposed method does not require multiple adjustments of cameras and projectors, nor does it necessitate the projection of a large number of fringes, as shown in Table 4. With low hardware costs, the image acquisition time in experiments could be controlled within 0.5 s, making it efficient. Compared with the methods using auxiliary devices, the proposed system boasted a simple structure and strong measurement stability.

Table 4 Comparison of number of fringe projection

Measurement method	Number of fringe projections
Rao's method[2]	20
Jiang's method[4]	72
Feng's method[6]	36
Lin's method[8]	31
Proposed method	12

4 Conclusions

To address the issue of missing point clouds when measuring the tongue surface with traditional binocular system, a trinocular-projection measurement system was designed and constructed based on fringe projection. The system consisted of a trinocular system and three binocular subsystems, relying on stereoscopic calibration to obtain the spatial relationship among the three cameras and unify the coordinate systems of multiple systems. In the stereo matching step of the trinocular system, dual-epipolar constraint was introduced based on phase and order constraints to improve matching accuracy. Leveraging the advantages of the trinocular system, the initial matching point set was supplemented to enhance point cloud density, completing the supplementation of missing point clouds in high-reflectance areas. Finally, experimental measurements of tongue models and real tongue surfaces demonstrated that the proposed system could obtain well-performing point clouds. Especially in the real tongue surface measurement, it could generate point clouds with clear texture and complete features, supplementing information on high-reflectance areas of the tongue surface. Further measurement experiments on other objects with high reflectivity verified the effectiveness of the proposed method.

References

[1] WANG Y H, ZHANG Q, HU Y, et al. 3D small-field surface imaging based on microscopic fringe projection profilometry: a review. Chinese Optics, 2021, 14(3): 447-457.

[2] RAO L, DA F P. High dynamic range 3D shape determination based on automatic exposure selection. Journal of Visual Communication and Image Representation, 2018, 50: 217-226.

[3] JIANG M N, LI F, YAN T T. 3D measurement of highly-reflective surface based on phase detection of projected grating. Electronics Optics and Control, 2021, 28(1): 94-97.

[4] JIANG H Z, ZHAO H J, LI X D. High dynamic range fringe acquisition: a novel 3-D scanning technique for high-reflective surfaces. Optics and Lasers in Engineering, 2012, 50(10): 1484-1493.

[5] PENG G Z, CHEN W J. Fringe pattern inpainting based on convolutional neural network denoising regularization. Acta Optica Sinica, 2020, 40(18): 1810002.
[6] FENG W, XU S N, WANG H H, et al. Three-dimensional measurement method of highly reflective surface based on per-pixel modulation. Chinese Optics, 2022, 15(3): 488-497.
[7] LIU F, HE C Q, SHEN A M, et al. Optimized compensation method of divisional projection for saturated region of structured light. Acta Optica Sinica, 2018, 38(6): 0612001.
[8] LIN H, GAO J, MEI Q, et al. Adaptive digital fringe projection technique for high dynamic range three-dimensional shape measurement. Optics Express, 2016, 24(7): 7703-7718.
[9] ZHAO S B, LIU L Y, MA M Y. Adaptive high-dynamic range three-dimensional shape measurement using DMD camera. IEEE Access, 2019, 7: 67934-67943.
[10] SALAHIEH B, CHEN Z Y, RODRIGUEZ J J, et al. Multi-polarization fringe projection imaging for high dynamic range objects. Optics Express, 2014, 22(8): 10064-10071.
[11] RIVIERE J, RESHETOUSKI I, FILIPI L, et al. Polarization imaging reflectometry in the wild. ACM Transactions on Graphics, 2017, 36(6): 206.
[12] FENG S J, CHEN Q, ZUO C, et al. Fast three-dimensional measurements for dynamic scenes with shiny surfaces. Optics Communications, 2017, 382: 18-27.
[13] LIU G H, LIU X Y, FENG Q Y. 3D shape measurement of objects with high dynamic range of surface reflectivity. Applied Optics, 2011, 50(23): 4557-4565.
[14] NAYAR S K, IKEUCHI K, KANADE T. Surface reflection: Physical and geometrical perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(7): 611-634.
[15] LI M H, CAO Y P, WU H T. Three-dimensional reconstruction for highly reflective diffuse object based on online measurement. Optics Communications, 2023, 533: 129276.
[16] MA Y P, LI Q W, XING J, et al. An intelligent object detection and measurement system based on trinocular vision. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(3): 711-724.
[17] CONEN N, LUHMANN T, MAAS H G. Development and evaluation of a miniature trinocular camera system for surgical measurement applications. Journal of Photogrammetry Remote Sensing & Geoinformation Science, 2017, 85(2): 127-138.
[18] XIAO Y, BIN L K. A prism-based single-lens stereovision system: From trinocular to multi-ocular. Image and Vision Computing, 2007, 25(11): 1725-1736.
[19] FU L H, WANG C Y, HE J J, et al. Camera pose measurement method based on feature matching. Journal of Measurement Science and Instrumentation, 2023, 14(1): 1-8.
[20] TOWERS C E, TOWERS D P, JONES J D C. Absolute fringe order calculation using optimised multi-frequency selection in full-field profilometry. Optics and Lasers in Engineering, 2005, 43(7): 788-800.
[21] ZHEN T X, HUANG S, LI Y F, et al. Key techniques for vision based 3D reconstruction: a review. Acta Automatica Sinica, 2020, 46(4): 631-652.
[22] STENTOUMIS C, GRAMMATIKOPOULOS L, KALISPERAKIS I, et al. A local adaptive approach for dense stereo matching in architectural scene reconstruction. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2013, XL-5/W1: 219-226.
[23] KAZHDAN M, HOPPE H. Screened poisson surface reconstruction. ACM Transactions on Graphics, 2013, 32(3): 29.

基于条纹投影三目系统的舌面完整点云获取方法

王　鹏*, 张　学, 孙长库, 付鲁华

天津大学 精密测试技术及仪器全国重点实验室, 天津 300072

摘　要: 传统的舌诊过程存在客观性较差的问题, 将计算机视觉技术应用于舌诊可以有效推进舌诊客观化。双目立体视觉结合结构光条纹投影技术是三维测量的常用手段, 但对于舌诊的测量场景, 由于舌体表面常带有津液和口水等, 造成条纹图像出现分布随机性较大的高反光区域, 导致相位解算错误和点云缺失。本文提出了一种基于光栅投影的三目测量系统, 由3个相机构成1个三目和3个双目子系统。在相位和级次约束基础上引入双极线约束, 提高三目立体匹配精度, 并利用匹配点补充优化三目匹配点集, 重建高反光区域点云。结果显示, 对比传统双目系统, 本系统的匹配精度和重建精度均得到提升, 特别是在真实舌面测量中能够生成纹理清晰、特征完整的点云, 对舌面高反光区域进行了有效测量, 有助于舌诊的客观化进行。

关键词: 三维测量; 三目系统; 光栅投影; 双极线约束

引用格式: WANG Peng, ZHANG Xue, SUN Changku, et al. A method for obtaining complete point cloud of tongue surface based on fringe projection trinocular system. Journal of Measurement Science and Instrumentation, 2024, 15(2): 166-175.