The Master Manipulator Control System in a Master-Slave Robot based on Trinocular Stereo Vision

Qiong Liu*, LeiLei Cao, Jiang Chen and ZhaoXia He

Changan University, Key Laboratory of the Ministry of Education On Road Works Technology & Equipment, Xian, ShaanXi, China

Abstract: In this paper, a master manipulator control system in a master-slave robot which uses trinocular stereo vision is presented. The system utilizes three real-time cameras to identify and track the movements of the master handle, while also measuring the master handle position. The system follows a basic four step process. First, within the image the master handle is identified using the contrast between the handle and the background. Second, the speed and direction of the master handle's next position is predicted in real-time by using the Kalman filter. Third, the position of the master handle is measured in real time using stereo matching, which utilizes a gray area matching technique to consistently identify the proper, unique image and eliminate spurious matching. This study concludes with the experimental results that demonstrate the feasibility of the master manipulator control system.

1 Introduction

The master-slave robot is an important development in the field of robotics, particularly because of its potential use and advantages in hazardous or difficult to access enviornments and landscapes. In general, a master-slave robot includes a master handle and slave robot. The master handle is the tool through which the operator exchanges information to the slave robot. Domestic and foreign scholars have proposed a variety of master handle configurations. Farrington and Gerling [1] proposed a Phantom-type master handle with a multi-dimensional force feedback which can simulate the forces in a virtual environment. Hung et al. proposed the 6-DOF master handle [2,3,4], which is based upon the parallel connection mechanism that is known for its fast motion in a larger workspace. Baser [5] proposed a new type of master high dynamics handle, based on the dual parallel connection mechanism, which utilizes a rigid drive and flexible drive. Berkelman proposed a non-contact magnetic levitation master handle that can achieve six degrees of freedom motion [6,7] and force feedback. Several of these master handle configurations have a complex mechanical structure which limits their functionality in a limited working space enviornment.

In recent years, there have been significant developments in visual detection technology. More specifically, vision detection has become faster and more accurate in its measurements, and become increasingly vital in tracking moving targets.. Gao studied human fingertip motion, which can be used for real-time tracking and recognition by the dexterous master-slave operation [8,9]. This fingertip motion tracking can be completed in 0.4ms when tracking multi-markers. Nash studied the tracking of objects moving through three-dimensional space [10], with a focus on what happens when the target moves beyond the field of view and the camera adjusts automatically. Visual detection has certain advantages over other methods of tracking beacuse it is non-contact, high speed, and highly flexible, —attributes that make it easier to measure a moving object through a large space. In this study, the movement of the master handle is identified and tracked by three cameras. The position of the master handle is measured by trinocular stereo vision. The position and orientation of the master handle is then sent to the slave robot. In this way, a simple mechanical structure that functions in a large workspace is achieved.

* Corresponding author e-mail: 18909290986@163.com

2 The Structure of the Master-Slave Robot System

The structure of the master-slave robot system is shown in Figure 1. There is a master handle subsystem and slave robot subsystem in the dashed box. The master handle subsystem consists of the operating handle, the visual detection unit, the computer control unit, and the monitor. The slave robot subsystem includes the robot, control appliance, and surveillance cameras, as well as other components.

The operator controls the movement of the operating handle. The position and posture of the operating handle can be identified, tracked and measured in real-time by the visual detection unit. Through the master-slave system's movement mapping function, the position and posture of the operating handle is converted into joint movement parameters and then the joint movement parameters are used for driving the robot as a corresponding movement of the operating handle. At the same time, in real time, the surveillance cameras register the position and pose of robot. Meanwhile, the operator makes a decision on the operating handle's next movement after reviewing the real-time video. Taken together, these operational steps complete the work cycle.
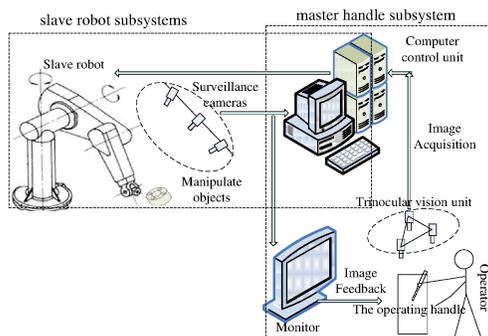

Fig. 1: The structure of Master-slave robot system

3 The design of the Master handle Subsystem

3.1 The Principle of the Master handle Subsystem

The main function of the master handle subsystem is to facilitate a human-computer interaction. The visual detection unit collects information on the operator motion, which is then used to instruct the motion of the robot through the computer control unit. The operator can move the operating handle into the next position using the real-time feedback of the moving robot. The working block diagram of the master handle subsystem is shown in Figure 2.
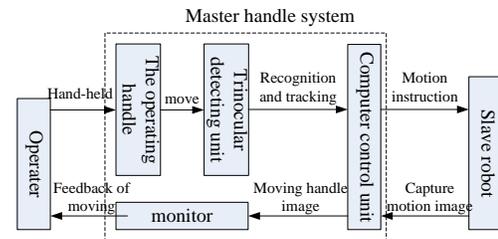

Fig. 2: the block diagram of the master handle subsystem

3.2 The work space of the master handle

The work space of the master handle includes the field-of-view of the multi-cameras. In order to avoid the arm of the operating handle being concealed, the handle should be tracked and located by the three cameras from three different perspectives, as shown in Figure 3.

Each camera oversees one field of view. Each of the three camera's field of views overlap with each other at certain angles, as shown in Figure 4. The work space of the operating handle is where the three field of views overlap. Therefore, the work space of the master handle is a 3D measurement with a strange shape and an unknown position and size [11,12]. It is difficult to translate this particular measurement.

It is also neccesary to calculate the location and size of the workspace. In order to simplify this process, the field of view for the cameras is expressed in conical space, as shown in Figure 5.

As shown in Figure 5, the baseline distance B is the line between the CCD1 and CCD2 lens center. The optical axis of camera CCD1 and CCD2 intersect at angle β_1 and β_2. The CCD1 lens center is viewed as the grid origin, the baseline B is viewed as the X axis; the Z axis is perpendicular to the baseline B on the plane defined by the two intersecting optical axes; and the Y axis direction is determined by the right-hand rule. The incident direction of the Camera CCD 3 is perpendicular to the XOZ plane.
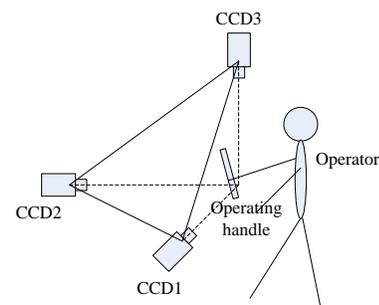

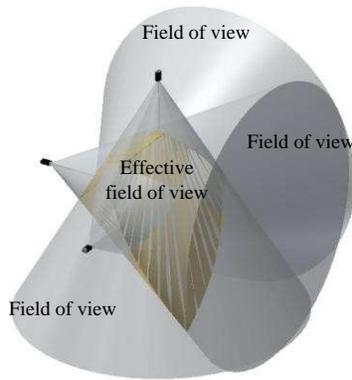**Fig. 3:** the multi-camera layout of the master handle subsystem

Fig. 4: the overlapping field of three camera views

Assume the original position of the conical space as shown in the figure to the right in figure 5. It is known that the Z-axis is the rotation axis and the conical surface parametric equation is $\Lambda : \begin{cases} x = a \cdot \cos t \\ y = a \cdot \sin t \\ z = a \cdot \cot \omega \end{cases}$ with alignment $x^2 + y^2 = a^2$, $z = a \cdot \cot \omega$, including $t \in R$, $\omega = \arctan(\sqrt{T_x^2 + T_y^2} / f)$.
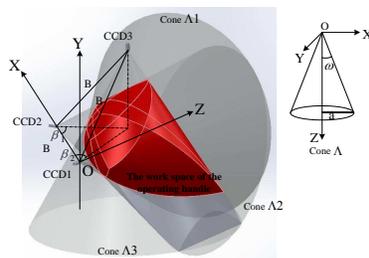

Fig. 5: the shape and position of the workspace

Then, the cone Λ rotates around Y-axis through a clockwise $(90° \text{-} \beta_1)$ degrees. It is understood, then, that in the coordinate system OXYZ the rotated factor matrix of CCD1 conical surface is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \cos(90° - \beta_1) & 0 & -\sin(90° - \beta_1) \\ 0 & 1 & 0 \\ \sin(90° - \beta_1) & 0 & \cos(90° - \beta_1) \end{bmatrix} \cdot \begin{pmatrix} a \cos t \\ a \sin t \\ a \cot \omega \end{pmatrix}$$

Therefore, the CCD1 conical surface parametric equation is:

$$\Lambda 1 : \begin{cases} x = a \sin \beta_1 \cos t - a \cos \beta_1 \cot \omega \\ y = a \sin t \\ z = a \cos \beta_1 \cos t + a \sin \beta_1 \cot \omega \end{cases} \quad (1)$$

Similarly, the cone Λ rotates around the Y-axis through a counter-clockwise $(90° \text{-} \beta_2)$ degrees, and shifts

B along the X-axis. Thus, it is understood that in the coordinate system OXYZ the rotated factor matrix of the CCD2 conical surface is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \cos(-(90° - \beta_2)) & 0 & -\sin(-(90° - \beta_2)) \\ 0 & 1 & 0 \\ \sin(-(90° - \beta_2)) & 0 & \cos(-(90° - \beta_2)) \end{bmatrix}$$
$$\cdot \begin{pmatrix} a \cos t \\ a \sin t \\ a \cot \omega \end{pmatrix} + \begin{pmatrix} B \\ 0 \\ 0 \end{pmatrix}$$

Therefore, the CCD2 conical surface parametric equation is:

$$\Lambda 2 : \begin{cases} x = a \sin \beta_2 \cos t + a \cos \beta_2 \cot \omega + B \\ y = a \sin t \\ z = -a \cos \beta_2 \cos t + a \sin \beta_2 \cot \omega \end{cases} \quad (2)$$

Then, the cone Λ rotates around X-axis through a clockwise $90°$ degrees, shifts B/2 along the X-axis, shifts B/2 along the Z-axis and shifts $B/\sqrt{2}$ along the Y-axis. It is understood, then, that in the coordinate system OXYZ the rotated factor matrix of the CCD3 conical surface is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \cos 90° & 0 & -\sin 90° \\ 0 & 1 & 0 \\ \sin 90° & 0 & \cos 90° \end{bmatrix} \cdot \begin{pmatrix} a \cos t \\ a \sin t \\ a \cot \omega \end{pmatrix} + \begin{pmatrix} B/2 \\ B/\sqrt{2} \\ B/2 \end{pmatrix}$$

Therefore, the CCD3 conical surface parametric equation is:

$$\Lambda 3 : \begin{cases} x = -a \cot \omega + B/2 \\ y = a \sin t + B/\sqrt{2} \\ z = a \cos t + B/2 \end{cases} \quad (3)$$

Assuming that the depth of field is in perfect condition for the three cameras CCD1, CCD2, CCD3, the effective field of view for the trinocular consists of a cluster of points (Ω) which lie in the overlapping field of the three camera views. The cluster of points Ω is expressed by the formula (1) to (3):

$$\Omega = \{(x,y,z)/(x,y,z) \subseteq (\Lambda 1 \cap \Lambda 2 \cap \Lambda 3)\} \quad (4)$$

To review, the workspace of the master handle is determined by the set of points in the overlap zone.

4 Recognition and tracking algorithm of the master handle

4.1 Recognition of moving target based on background subtraction

The background subtraction method is an efficient recognition algorithm used to track a moving object. In

general this process begins by selecting one or several images as a background template and then subtracting in the chosen image from the background template by using this formula (5). If the number of nonzero pixels in the differential image is greater than the preset threshold value, it shows a moving object in the image. Let $f_k(x,y,t)$ and $f_b(x,y,t)$ denote the current frame image at time t and the background template separately, so that the differential image D $_k$ is:

$$D_k(x,y,t) = \begin{cases} f_k(x,y,t) & |f_k(x,y,t) - f_b(x,y,t)| > T \\ 0 & others \end{cases} \quad (5)$$

In each of these formulas, T is the set threshold value; the nonzero region in D $_k$ is the region of the moving target. Because the background subtraction method can be influenced by changes in the environment, light intensity, and other factors, it is necessary to promptly update the background templates in an effort to nullify these influential effects. The background-updating algorithm [13,14,15] proposed by Lamard is:

$$B_k(x,y) = \alpha f_k(x,y) + (1-\alpha)B_{k-1}(x,y) \quad (6)$$

Wherein α is the updating weights of the background templates that is used for adjusting the updating rate of the background templates.

4.2 Motion Tracking Based on Kalman Filter

After using background subtraction to determning the moving target, a maximum bounding rectangle W_{ROI} of the target contour is used as the window for the motion tracking. Then, in order to narrow the moving target search area, the next location of the moving target is predicted with the Kalman filter, according to the current status (position, velocity, acceleration) of the moving target.

Suppose the status vector of the Kalman filter is $X = (x,y,v_x,v_y,a_x,a_y)^T$, which separately denotes position, velocity and acceleration for the x, y-direction in the image; and the observation vector is $Z = (x,y)^T$. The prediction algorithm is shown, then, as the following steps:

(1) Initialization: the model of the Kalman filter is:
State equation:

$$X_{k+1} = \Phi_k X_k + \omega_{k+1}, \ \omega_{k+1} \in N(0,Q) \quad (7)$$

where Φ_k is a state transition matrix; ω_{k+1} is a random interference vector.
Observation equation:

$$Z_{k+1} = H_k X_k + \upsilon_{k+1}, \ \upsilon_{k+1} \in N(0,R) \quad (8)$$

where H_k is an observation matrix; v_{k+1} is a observation noise vector.

The Kalman filter should be initialized when it is used for the first time. The detected initial state X_0 of the moving target is assigned to the Kalman filter as the initial value. If the velocity and acceleration of the moving target is unknown, the assignment is zero. Meanwhile, the time of the current image is recorded and the covariance of the initial error is set to zero.

(2) The prediction of moving target:
Status prediction equation:

$$\widehat{X}_{k+1} = \Phi_k X_k + \omega_{k+1} \quad (9)$$

Error covariance prediction equation:

$$P_{k+1}^- = \Phi_k P_k \Phi_k + Q \quad (10)$$

$$\Phi_k = \begin{bmatrix} I & \Delta t \cdot I & \frac{1}{2}\Delta t^2 \cdot I \\ 0 & I & \Delta t \cdot I \\ 0 & 0 & I \end{bmatrix} \quad (11)$$

Where, Δt is the sampling time, Φ_k is a 6×6 state transition matrix, I is a 2×2 unit matrix.

(3) Update process:
Observation matrix:

$$H_{k+1} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (12)$$

Kalman gain coefficient:

$$K_{k+1} = P_{k+1}^- H_{k+1}^T (H_{k+1} P_{k+1}^- H_{k+1}^T + R)^{-1} \quad (13)$$

Covariance update equation:

$$P_{k+1} = (1 - K_{k+1} H_{k+1}) P_k^- \quad (14)$$

Status update equation:

$$X_{k+1} = \widehat{X}_{k+1} + K_{k+1}(Z_{k+1} - H_{k+1}\widehat{X}_{k+1}) \quad (15)$$

5 Gesture measurement algorithm

The gesture of the operating handle can be determined by measuring the coordinates of certain marks on the handle.

The measuring principle is such that while the operator is moving certain markers on the handle will be concealed. To accomdate this principal, the trinocular detecting unit can calculate the handle gesture as long as two of the cameras are able to detect four or more markers. As such, the measuring algorithm is divided into three steps: 1) recognition and location of the markers; 2) marker matching; and 3) gesture calculation. Here's an example in which any two cameras can be used for the algorithm.

5.1 Recognition and location of markers

The binarization image processed by the iterative method: the number of the markers N_0 (a known quantity) is viewed as a condition for the ending iteration. The choice of the threshold is:

$$T_0 = (T_{\max} + T_{\min})/2 \qquad (16)$$

where $T_{\max}$, $T_{\min}$ are separately the initial maximum, minimum threshold value. Assume that the max number of markers detected is N. In order to reduce marker loss, it should be that $N = \lambda N_0$, λ is greater than 1. First, the image is binarized by the threshold T_0; and then the contour number n is detected, if $n > N$, it takes $T_{\min} = T_0$; if $n < N$, it takes $T_{\max} = T_0$; if $n = N$, it ends the iteration and the final T_0 is the selected threshold.

After the binarization processing, the contour extraction will begin. Because the markers are round, and the marker's size and the relative position are known quantities, the pseudo markers in the image can be removed according to the shape of the marker's contour, the size, and the position. With this information the real markers can be determined. The real markers are extracted at the sub-pixel level by the centroid method [16,17,18]; the image coordinates of the centroid for each marker are denoted by $\{P_{Li}|i = 1, 2, \cdots N_0\}$, $\{P_{Rj}|i = 1, 2, \cdots N_0\}$.

5.2 The matching of the markers

During the marker matching process, the marker centroids are taken as the matching primitive of the feature, and the relevant neighborhood windows of the centroids are taken as the matching primitive of the gray area. As such the matching process can be achieved by using the complex primitive.

Suppose the relevant neighborhood windows of the centroids are denoted by W; the search area is W_{ROI} (which was determined in the previous section); and the matching algorithm is as follows:

(1) The initial matching of the feature points is taken by the epipolar constraint and then the candidate points are obtained. Suppose P_i, P_j represent separately the corresponding point of the feature in the source image and the target image. L_i represents the corresponding epipolar of P_i in the target image. D_{ij} represents the distance between P_j and L_i. A_i represents a cluster of the candidate points for P_i.

$$D_{ij} = f(P_i, P_j, L_i) \qquad (17)$$

Suppose the distance threshold value is D_0, if $D_{ij} < D_0$, then P_j will be added to the cluster A_i.

(2) Because of the similarity of the gray value within the neighborhood windows around the feature points, the matching might not be fully accurate. This can be described the by the normalized cross-correlation coefficient [19]:

$$S(x,y) = \frac{\sum\limits_{y'=0}^{n-1}\sum\limits_{x'=0}^{m-1}(T(x',y')-\bar{T})(I(x+x',y+y')-\bar{I})}{\sqrt{\sum\limits_{y'=0}^{n-1}\sum\limits_{x'=0}^{m-1}(T(x',y')-\bar{T})^2\sum\limits_{y'=0}^{n-1}\sum\limits_{x'=0}^{m-1}(I(x+x',y+y')-\bar{I})^2}}$$

$$(18)$$

Where, $S(x,y)$ is the similarity of two $m \times n$ corresponding neighboring windows, I(x, y), T(x, y) are the gray value in the target image, and the source image at coordinate (x, y), $\bar{I}$ and $\bar{T}$ are the gray average within the neighborhood window in the target image and the source image. Suppose the similarity threshold is ε_0. If the similarity of the point-to-point (P_{Li}, P_{Rj}) is $S_{ij} > \varepsilon_0$, the candidate points are retained or excluded.

(3) Sequential consistency constraints and unique constraints are introduced to eliminate spurious matches. In the above matching process, a match must be bidirectional, which means that a search of the matching points must be carried out between the left and right image. The corresponding point P_R of left imaging point P_L is searched for in the right image; in reverse, the corresponding point P_L of right imaging point P_R is searched for in the left image. If the bidirectional match is not a success, the match itself is incorrect and this pair of corresponding points should be excluded.

After the above steps have been completed, the best matching points of each marker can be accurately found.

5.3 The gesture calculation of the operating handle

The gesture calculation is divided into three steps: 1) 3D coordinates of each marker are found in any two of the trinocular system cameras. The experiment shows that if the arm conceals one marker then that marker's coordinates can be calculated by one pair of cameras in the binocular system; if the marker isn't concealed then the marker coordinates can be calculated using three of the binocular system cameras. 2) The optimal coordinate values of the markers need to be determined. This information, in turn, will determine the unique gestures of the operating handle. Because the coordinates of a marker will be different in each of the binocular system cameras, the coordinates need to be transformed into the same coordinate system. The several coordinates of a marker are then averaged to determine its optimal coordinates. 3) According to the structural parameters of the operating handle and the 3D coordinates of its markers, the gesture of the operating handle can be solved. The algorithm is as follows:

First, suppose the world coordinate system and the CCD1 camera coordinate system coincide with each other in the trinocular system. The internal and external

parameters for the three groups of the trinocular system are separately obtained through calibration, including the rotation matrix R_{rli}, translation vector T_{rli} and effective focal length f_{li}, f_{ri} $(i = 1, 2, 3)$, when i represents the group number of the binocular system.

$$\begin{cases} R_{rli} = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix} \\ T_{rli} = \begin{pmatrix} t_x & t_y & t_z \end{pmatrix}^T \end{cases} \quad (19)$$

$p_l(X_l, Y_l)$ and $p_r(X_r, Y_r)$ are the image coordinates of the marker P in the left and right image plane. According to the relation of perspective projection for the marker and its corresponding points in the left and right image plane, the 3D coordinates $P(x, y, z)$ in the CCD1 camera coordinate system are:

$$\begin{cases} x = zX_l/f_{li} \\ y = zY_l/f_{li} \\ z = \dfrac{f_{li}(f_{ri}t_x - X_r t_z)}{X_r(r_7 X_l + r_8 Y_l + r_9 f_{li}) - f_{ri}(r_1 X_l + r_2 Y_l + r_3 f_{li})} \\ = \dfrac{f_{li}(f_{ri}t_y - Y_r t_z)}{Y_r(r_7 X_l + r_8 Y_l + r_9 f_{li}) - f_{ri}(r_4 X_l + r_5 Y_l + r_6 f_{li})} \end{cases}$$

$$(20)$$

Second, as shown in the trinocular system in Figure 5: the two groups of the binocular system, CCD1-CCD2 and CCD1-CCD3, are used to determine the 3D coordinates of the marker. When using the third group of the binocular system, CCD2-CCD3, the 3D coordinate value $P^3(x_3, y_3, z_3)$ of the marker needs to coordinate a conversion to the CCD1 camera coordinate system. The rotation matrix R_{rl1}, translation vector T_{rl1} can be obtained through calibration; the equation of the coordinate conversion is:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = R_{rl1} \begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix} + T_{rl1} \quad (21)$$

where (x_1, y_1, z_1) is the coordinate value of the point $P^3(x_3, y_3, z_3)$ in the CCD1 camera coordinate system.

Third, the marker coordinates are converted through a coordinate conversion, but only after they have been determined by each binocular system. Then, several coordinate values of the same marker are averaged out and the optimal coordinate value is determined. The marker P coordinates in the CCD1 camera coordinate system are:

$$P^W = P^{CCD1} = (\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i, \frac{1}{n}\sum_{i=1}^{n} z_i), n = 1, 2, 3.$$

$$(22)$$

Fourth, the coordinate matrix $M = (P_1^W, P_2^W, ..., P_i^W, ..., P_n^W)$ is composed of the 3D coordinates of all the markers. Where $P_i^W = x_i, y_i, z_i)^T$, $i = 1 ... n$, n is the number of markers.

Finally, a coordinate system O-xyz is built for the operating handle, in which the position and orientation of the handle is indicated by ^wF,

$$^wF = \begin{pmatrix} n_x & o_x & a_x & x_O \\ n_y & o_y & a_y & y_O \\ n_z & o_z & a_z & z_O \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (23)$$

Where $(n_x, n_y, n_z)^T$, $(o_x, o_y, o_z)^T$ and $(a_x, a_y, a_z)^T$ is the unit direction vector, which separately represents the orientation of three axes for the coordinate system O-xyz in the world coordinate system. $(x_o, y_o, z_o)^T$ is the coordinate value of the origin O in the world coordinate system.

According to the structure parameters of the handle, the coordinate value of each marker in the coordinate system O-xyz can be calculated, which is show as $P_i' = (x_i', y_i', z_i')^T$, $i = 1 ... n$, n is the number of markers. Then, the coordinate matrix $M' = (P_1', P_2', ..., P_i', ..., P_n')$ is composed of 3D coordinates of all the markers in the coordinate system O-xyz, therefore,

$$M = {}^wF \cdot M' \quad (24)$$

Substituting M and M', the gesture of the handle can be expressed by ^wF.

The attitude angle of the handle can be expressed by the Euler angles Z-Y-X [20]. Therefore, the gesture of the handle ^wF is the vector $((x, y, z), \alpha, \beta, \gamma)$, where (x, y, z) represents the position of the handle and α, β, γ is the angle which separately represents the rotations for the handles relative to the world coordinate system's Z-axis, Y-axis, X-axis.

6 Experimental analysis

6.1 The structure of experimental system

The hardware for the experimental system consists of the following devices. Three industrial cameras, GRAS-20S4M-C with 2million pixels and a maximum frame rate of 30 frames/s;. The camera lens are Kowa LM12HC with a 12.5 mm focal length; the computer configuration is Intel Xeon X3220 2.4GHz and 2GB RAM; a precise 3D motion platform whose positioning accuracy is 0.02mm with a distance of 200mm×200mm×150mm. Also, a binocular system was used to carry out the experimental verification of the vision measurement of the handle gesture.

6.2 The camera calibration of the binocular vision system

A familiar calibration method was adopted to calibrate two of the cameras of the binocular system separately.

These calibrations yielded the camera's Internal and external parameter [21, 22].

Left camera CCD1 internal parameter matrix:

$$A_l = \begin{pmatrix} 2956.2067 & 0 & 627.9930 \\ 0 & 2940.0975 & 515.2024 \\ 0 & 0 & 1.0000 \end{pmatrix}$$

Right camera CCD2 internal parameter matrix:

$$A_r = \begin{pmatrix} 2939.5353 & 0 & 667.5260 \\ 0 & 2922.2544 & 502.6779 \\ 0 & 0 & 1.0000 \end{pmatrix}$$

Suppose R_l, T_l or R_r, T_r separately represent the corresponding position relation between the left or right camera coordinate system and the world coordinate system. Then, the position relation between the left and right camera coordinate system can be calculated by the mathematic expression below.

$$\begin{cases} R_{rl} = R_l R_r^{-1} \\ T_{rl} = T_l - R_l R_r^{-1} T_r \end{cases} \tag{25}$$

The rotation matrix R_{rl} and translation vector T_{rl} can be obtained by the iterative optimization of 14 positions parameters for calibration target, namely,

$$R_{rl} = \begin{bmatrix} 0.9838 & -0.0045 & -0.1799 \\ 0.0092 & 1.0000 & 0.0224 \\ 0.1801 & -0.0241 & 0.9837 \end{bmatrix},$$

$$T_{rl} = \begin{bmatrix} 98.7723 & 0.8410 & 9.1032 \end{bmatrix}$$

6.3 Recognition and tracking of the handle

The length of the cylindrical handle used in the experiment is 200mm; the diameter is 10mm. Six columns markers are distributed evenly along the circle of the cylindrical surface atop the handle, and the angle of the two adjoining markers along the circle is 60°. Every column has two markers. Figure 6 shows part of the processing result of the identification and tracking for the handle. Figure 6(a) and (b) are the identification results of the handle after using the background difference method, while 6(c) and (d) are the tracking process of the handle after using the Kalman filter. According to the figures, the algorithm can track the motion of the handle precisely. The result shows that the algorithm introduced in this paper can adapt well to changing ambient light.

6.4 Gesture measurement of the handle

Precise 3D motion platform is used for driving the handle move within the workspace of the handle in the
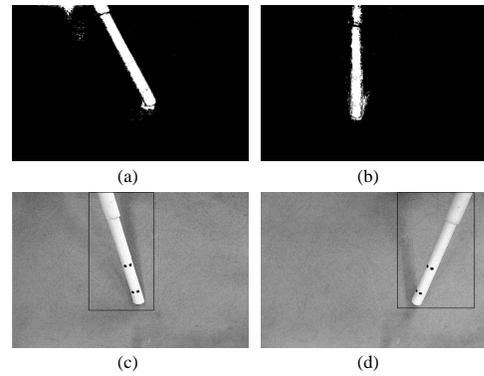

Fig. 6: Results for the recognition and tracking of the handle

experiment. The handle is fixed on the motion platform according to the setting attitude angle, and then the parameters of reality gesture for the handle can be obtained, and the reality parameters of handle gesture are used as theoretical value. As shown in Table 1. Meanwhile, the handle gesture is measured by binocular system. Compare the parameters measured with the theoretical parameters, visual measure precision can be analyzed. Table 1 provide part of the experiment result, from the data in the table, the algorithm can measure the handle gesture precisely.

6.5 Error analysis

The mean absolute error is introduced to represent the measurement precision of algorithm, and carry out quantitative analysis of the error. The mean absolute error of position measurement can be defined as:

$$E_t = \frac{\sum_{i=1}^{N} \sqrt{(x'_i - x_i)^2 + (y'_i - y_i)^2 + (z'_i - z_i)^2}}{N} \tag{26}$$

The mean absolute error of attitude angle can be defined as:

$$E_\theta = \frac{\sum_{i=1}^{N} |\theta_i' - \theta_i|}{N} \tag{27}$$

where (x, y, z) is the reality 3D coordinate, while (x', y', z') is the measured 3D coordinate. And $\theta' = (\alpha', \beta', \gamma')$ is the measured attitude angle, while $\theta = (\alpha, \beta, \gamma)$ is the reality attitude angle. N is the number of sample, $N = 20$. Through counting the data in the table1, the absolute error of the position measurement can be obtained: $E_t = 0.31$mm; and the absolute error of the angle measurement: $E_\alpha = 0.67°$, $E_\beta = 0.56°$, $E_\gamma = 0.65°$. So the measurement error can be influenced by the factors as lens marginal distortion, the measuring distance, the size of markers, the motion speed of handle, and these problems need to further resolve in the next research.

Table 1: Table 1 Result of handle gesture measurement

Series No.	Measured position coordinates(x', y', z')/mm	Theoretic position coordinates(x, y, z)/mm	Measured attitude angle$(\alpha', \beta', \gamma')$/°	Theoretic attitude angle$(\alpha', \beta', \gamma')$/°
1	(-114.59,36.91,598.83)	(-114.80,36.81,598.66)	(27.1, 77.3,-88.2)	(26.4,76.7,-89.1)
2	(-90.78,41.93,586.25)	(-90.57,41.90, 586.25)	(-123.0,77.1,-100.1)	(-123.6,76.8,-101.3)
3	(-77.54,45.73,587.22)	(-77.61,45.67,587.04)	(-102.9,67.0,-111.9)	(-103.6,66.1,-112.2)
4	(-45.79,48.69,578.99)	(-45.78,48.88,578.83)	(-32.1,81.2,92.6)	(-33.1,80.5,91.9)
5	(-21.47,50.42,572.92)	(-21.22,50.60,572.65)	(-117.2,62.1,-169.2)	(-118.2,61.8,-169.5)
6	(-7.80,45.64,569.41)	(-7.74,45.64,569.16)	(-95.2,68.7,-175.1)	(-95.7,68.2,-176.0)
7	(13.65,47.37,565.65)	(13.68,47.29,565.45)	(-85.4,68.8,-155.3)	(-86.3,68.4,-156.1)
8	(37.06,48.25,561.22)	(37.27,48.4289,561.55)	(-97.7,39.2,-178.5)	(-98.3,38.5,-179.0)
9	(45.70,42.58,559.16)	(46.09,42.41,559.35)	(-103.2,55.3,-167.6)	(-103.7,55.4,-168.1)
10	(62.13,45.91,557.43)	(62.31,45.88,557.29)	(-92.5,37.4,-148.6)	(-93.3,36.9,-149.0)
11	(74.48,47.47,557.00)	(74.48,46.92,557.19)	(-83.1,49.7,-177.8)	(-84.0,49.4,-178.7)
12	(70.24,3.21,555.88)	(70.58,3.13, 556.00)	(-87.2,50.3,-177.9)	(-87.7,49.4,-178.5)
13	(51.10,-0.04,558.07)	(50.84,0.22,557.93)	(-103.7,-54.5,-0.1)	(-103.9,-55.3,-0.7)
14	(43.38,-5.04,561.15)	(43.50,-5.12,561.14)	(-113.9,-53.2,-3.4)	(-114.5,-53.9,-3.7)
15	(31.93,-7.34,560.10)	(32.34,-7.43,559.88)	(89.2,-53.9,-169.3)	(88.3,-54.5,-170.1)
16	(16.70,-5.80,565.82)	(16.67,-5.69,566.08)	(-82.3,-44.4,4.6)	(-83.0,-44.2, 3.4)
17	(-22.15,-16.04,572.72)	(-22.44,-16.08,572.62)	(-107.5,-56.2,-0.1)	(-108.7,-56.9,-0.8)
18	(-39.50,-14.54,577.32)	(-39.51,-14.42,577.64)	(-49.6,45.7,178.6)	(-50.1,45.2,178.3)
19	(-69.72,-21.95,568.43)	(-69.75,-21.95,568.71)	(126.2,44.8,-1.2)	(125.2,44.1,-1.9)
20	(-84.07,-21.32,549.06)	(-84.06,-21.40, 549.00)	(-56.8,72.0,-172.4)	(-57.2,71.4,-172.6)

7 Conclusions

In order to solve the problem of the handle concealed, the moving handle can be recognized and tracked by three groups of binocular system in this paper. By the background difference method, the moving handle can be found out in image. For introducing the renew strategy, the influence of the change of the light to the detecting result is decreased. Then, through Kalman filter, the position, speed and acceleration of the handle can be forecasted, which narrow the hunting zone of target effectively and improve the detecting speed. For the stereo matching of markers, the methods of feature matching and regional gray matching are combined. By introducing the uniqueness and sequential consistency restriction, the pseudo matching can be rejected. At last, the spatial gesture of handle can be obtained by the three dimensional coordinates of marker points.

The experiment prove that the method presented in the paper can identify in real time and track the handle, measure its spatial gesture, and the measurement precision is 0.31mm, while the measurement precision of attitude angle is 0.67°.However, the measurement precision of the system and the robustness of the algorithm require to be improved, besides, how to introduce force feedback into the main operation handle need to do further study.

References

[1] Farrington, Mark W.; Gerling, Gregory J.; Kohan, Lynn R.; Johnson, Sheryl L. A multisensory illusion with haptic interaction to treat phantom limb pain. Haptics Symposium 2012, p 217-222, March 4-7, 2012. Vancouver, BC, Canada

[2] Hung, Vu Minh; Na, Uhn Joo. Tele-operation of a 6-DOF serial robot using a new 6-DOF haptic interface[J].2010 9th IEEE International Symposium on Haptic Audio-Visual Environments and Games. p 101-106, October 16 -17, 2010. Phoenix, AZ, United states

[3] Chotiprayanakul, P.; Liu, D. K.; Dissanayake, G. Human-robot-environment interaction interface for robotic grit-blasting of complex steel bridges. Automation In Construction. 2012, **27**, 11-23.

[4] Minh Hung Vu; Uhn Joo Na. A New 6-DOF Haptic Device for Teleoperation of 6-DOF Serial Robots. IEEE Transactions On Instrumentation And Measurement. 2011, **60**, 3510-3523.

[5] Baser, Ozgur; Konukseven, E. Ilhan. Kinematic calibration of PHANTOM Premium 1.5/6DOF haptic device. Key Engineering Materials, Advanced Design and Manufacture IV, 2011, **486**: 205-208

[6] Berkelman, Peter; Dzadovsky, Michae. Actuation model for control of a long range Lorentz force magnetic levitation device. 23rd IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010. p 4668-4673, October 18-22, 2010. Taipei, Taiwan

[7] Gumprecht, Jan D. J.; Lueth, Tim C.; Khamesee, Mir Behrad. Navigation of a robotic capsule endoscope with a novel ultrasound tracking system. Microsystem Technologies-Micro-And- Nanosystems-Information Storage And Procedure Systems. 2013, **19**, 1415-1423.

[8] Gao, Yuanqian; Li, Jinhua; Su, He; Li, Jianmin. Development of a teleoperation system based on virtual environment. 2011 IEEE International Conference on Robotics and Biomimetics, ROBIO 2011. p 766-771. December 7-11, Phuket, Thailand

[9] Buss, M.; Peer, A.; Schauss, T.; et al. Development of a Multi-modal Multi-user Telepresence and Teleaction System. International Journal Of Robotics Research. 2010, **29**, 1298-1316.

[10] Nash, James; Atanassov, Kalin; Goma, Sergio; Ramachandra, Vikas; Siddiqui, Hasib. Digital ruler: Real-time object tracking and dimension measurement using stereo cameras. Real-Time Image and Video Processing 2013. **8656**, February 6-7, Burlingame, CA, United states

[11] LIU Qiong, QIN Xiansheng, YIN Shenshun, HE Feng. Structural Parameters Design and Accuracy Analysis of Binocular Vision Measure System. China Mechanical Engineering. 2008, **19**, 2728-2730

[12] Garcia-Lazaro, Santiago; Albarran-Diego, Cesar; Ferrer-Blasco, Teresa; et al. Visual performance comparison between contact lens-based pinhole and simultaneous vision contact lenses. Clinical And Experimental Optometry. 2013, **96**, 46-52.

[13] Lamard, Laetitia; Chapuis, Roland; Boyer, Jean-Philippe. Dealing with occlusions with multi targets tracking algorithms for the real road context. 2012 IEEE Intelligent Vehicles Symposium, p 371-376, June 3-7, 2012. Alcal de Henares, Madrid, Spain

[14] Koppen, W. P.; Worring, M. Backtracking: Retrospective multi-target tracking. Computer Vision And Image Understanding. 2012, **112**, 967-980

[15] Schikora, Marek; Koch, Wolfgang; Cremers, Daniel. Multi-Object Tracking Via High Accuracy Optical Flow And Finite Set Statistics. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Prague Congress Ctr, Prague, CZECH REPUBLIC. MAY 22-27, 2011,1409-1412.

[16] Sandoz, Patrick; Elhechmi, Imen; Gharbi, Tijani. Toward stereo visual monitoring of three-dimensional translations with submicrometer resolution. Journal of the Optical Society of America A: Optics and Image Science, and Vision, 2012, **29**, 2451-2458

[17] Sedaghat, Amin; Ebadi, Hamid; Mokhtarzade, Mehdi. Image Matching of Satellite Data Based on Quadrilateral Control Networks. Photogrammetric Record. 2012, **27**, 423-442.

[18] Petrou, Maria; Jaward, Mohamed H.; Chen, Shengyong; et al. Super-resolution in practice: the complete pipeline from image capture to super-resolved subimage creation using a novel frame selection method. Machine Vision And Applications. 2012, **23**, 441-459.

[19] Zhang Guangjun. Machine Vision[M]. BeiJing: Science publishing company, 2005.

[20] Chao translation. The theoretics of robot [M]. BeiJing: the publishing company of mechanism industry, 2006.

[21] Xie, Meng; Wei, Zhenzhong; Zhang, Guangjun; Wei, Xinguo. A flexible technique for calibrating relative position and orientation of two cameras with no-overlapping FOV. Measurement: Journal of the International Measurement Confederation, 2013, **46**, 34-44

[22] Sargeant, Ben; Hosseininaveh, Ali A.; Erfani, Tohid; et al. A webcam photogrammetric method for robot calibration. Conference on Videometrics, Range Imaging, and Applications XII. Proceedings of SPIE. Munich, Germany, May 14-16, 2013, **8791**, 10.1117/12.2020484

Qiong Liu, PhD, Lecturer, she graduate from Northwestern Polytechnical University and major in machine vision and advanced manufacturing technology.

Leilei Cao, PhD, Associate Professor, she graduate from Northwestern Polytechnical University and major in the finite element analysis, the meshless numerical algorithm.

Jiang Chen, PhD, Lecturer, he graduate from Northwestern Polytechnical University and major in Mechanical and Electronic Engineering.

Zhaoxia He, PhD, Lecturer, she graduate from Northwestern Polytechnical University and major in Mechanical manufacturing and automation.