Geometric Calibration of Rotational Vision System for Dynamic Exterior Orientation

YAN Guangzong, ZHANG Jin*, CHENG Zhuo, Qiu Mouzhe,
HONG Lu, LI Weishi, ZHANG Rui, XIA Haojie

(*Anhui Province Key Laboratory of Measuring Theory and Precision Instrument*, *School of Instrument Science and Opto-Electronics Engineering*, *Hefei University of Technology*, *Hefei 230009*, *China*)

Abstract: Rotational Vision System (RVS) is a common active vision system with only rotational degrees of freedom. Usually, the degree of freedom for rotation is provided by the turntable and pan head. Or the hand to eye (EIH) structure in articulated arm robots. Due to assembly deviations and manufacturing accuracy limitations, the ideal assumption that the rotation axis is fully aligned with the coordinate axis of the local camera is mostly violated. To address this issue, we propose a generalized deviation model that specifies a rotation axis that connects the rotational motion of the platform with the external orientation (EO) of the camera. On this basis, we propose a heuristic estimation algorithm to minimize global reprojection errors and fit circles in space under constraints of global optimization. The experiment shows that the translation and tilt average reprojection errors of dynamic EO reconstruction based on the reprojection error method are 0.14 and 0.08 pixels, respectively. In the absence of angle measurement, the results of the circle fitting method are similar to them (with a relative error of about 2%), meeting the application requirements of general visual measurement.

Keywords: Active Vision, Pan-Tilt-Zoom, Eye-In-Hand, Calibration, Constrained Global Optimization

1 Introduction

RVS has been used in PTZ cameras [1,2], UAV aerial photography[3], and robot manipulators [4] for applica-tions such as visual tracking[5,6] and contactless measurement. However, there is a practical issue that has been addressed in literature [7-11]: due to the virtuality of rotation and optical axis, it's not possible to align the coordinate axes completely. This affects the reliability and accuracy of tracking and measuring. Davis and Chen [10] used pan-tilt parameters in the imaging process for pan-tilt camera motion and constructed a virtual calib-ration object using the wide area surveillance system. Byun et al. [9] proposed an alternative approach based on an independent pan/tilt axis model, which calibrates the misalignment axis relationship by fitting simulated circular trajectories created by corner points of a planar checkerboard.They also formulated the inverse kine-matics for servo control using 3D point transfor-mations. Li et al. [12] used a nonlinear parameter mapping to correct the camera rotation angle, obtaining pan and tilt angles via a quadratic equation of tangent. However, their single-point-calibration-method assumes parallel rotation axes with the world XY axes, affecting accuracy and generality. Wu and Radke [11] compensated for mechanical errors by modeling the camera rotation angle as linear. However, their method may result in propagation errors for large working distances. The precise transformation of the coordinate

system between the camera and the world is crucial for 3D vision applications. Although this task is known by different names in different domains, the approach involves establishing feature matching. However, there is a slight translation when the camera is rotated and obtaining robust estimates is challenging. To address this problem, some work has focused on relative pose estimation using camera networks. Nagayoshi and Pollefeys [13] proposed a circle-pose estimation method to estimate relative camera poses from circular trajectories generated by the pan-tilt rotation of a marker set on a camera. Chen et al. [14] developed the ITPC method for calculating head-eye parameters of a robotic bionic stereo system using a dual quaternion-based approach to improve accuracy. Rebello et al.[15]studied DCC calibration for external transfor-mations between cameras in active visual SLAM, using a pose-loop error optimization algorithm to enhance accuracy. Chen et al. [16] estimated the rotation axis direction and circle centers of closed circle trajectories from 3D corner points using a separate rotating actuated mechanism and camera, under specific assumptions. This paper has two main contributions. First, it proposes a general framework to model the assembly error between the camera and the rotating actuation platform, using the corresponding axes to reconstruct the external localization of the RVS with dynamic rotation angle information. Second, it proposes a way to estimate the rotation axis by minimizing the multi-view reprojection error and introduces a new 3D circle fitting algorithm. Experimental results show the effectiveness of these methods for 3D visual scanning of large targets, servo control, and motion target tracking with motion decoupling. These contributions have further research and application potential in various fields.

2 Model

Focusing on the accurate dynamic EO of camera, we propose a generalized mis-alignment axis-centre model for RVS which is illustrated in Fig.1. We take thesame idea with Davis and Chen [10] that rotations are occurring around arbitraryaxes in space so they can be regarded as spatially equivalent. For a certain rotationaxis we specify a corresponding axis attached to the camera for they are relativelycoincident in most cases.The rotation axis is abstractly represented by a normalized vector $\boldsymbol{n}$ and a point $\boldsymbol{O}$ on it with respect to the optical centre of the camera. Thispoint can be arbitrary but we choose the one that is a perpendicular intersectionfrom the optical centre to the rotation axis for simplicity. The geometric character-ization of a camera before and after it rotates in 3D space is denoted as $\boldsymbol{C}$-$\boldsymbol{X_cY_cZ_c}$ and $\boldsymbol{C'}$-$\boldsymbol{X'_cY'_cZ'_c}$. Similarly, the referenced world coordinate system is denoted by $\boldsymbol{W}$-$\boldsymbol{X_wY_wZ_w}$ for the sake of universality even though sometimes it can be replacedby that of the initial camera. Thus, all coordinates in the following discussion areunified in this referenced coordinate system. Given the camera proceeds from $\boldsymbol{C}$ to $\boldsymbol{C'}$ under rotation angle $\boldsymbol{\theta}$ of its platform (say, a pan-tilt turntable), the relativeorientation of it can be analytically gotten with Rodriguez formulation[17].

$$R_{c'c} = I + \sin\theta[\mathbf{n}]_\times + (1-\cos\theta)[\mathbf{n}]_\times^2 \tag{1}$$

It is quite simple a transform that

$$\begin{aligned} \overrightarrow{\mathbf{WC}} &= -R_{cw}^{\mathrm{T}}\mathbf{t}_{cw} \\ \overrightarrow{\mathbf{WC'}} &= -R_{c'w}^{\mathrm{T}}\mathbf{t}_{c'w} \end{aligned} \tag{2}$$

and the vector operation

$$\begin{aligned} \overrightarrow{\mathrm{OC}} &= \overrightarrow{\mathrm{WC}} - \overrightarrow{\mathrm{WO}} \\ \overrightarrow{\mathrm{OC'}} &= \overrightarrow{\mathrm{WC'}} - \overrightarrow{\mathrm{WO}} \\ \overrightarrow{\mathrm{CC'}} &= \overrightarrow{\mathrm{OC'}} - \overrightarrow{\mathrm{OC}}. \end{aligned} \tag{3}$$

After a short derivation process we get

$$\overrightarrow{\mathbf{OC'}} = \left(\cos\theta I + \sin\theta[\mathbf{n}]_\times\right)\overrightarrow{\mathrm{OC}} \tag{4}$$

Similar to (2), the absolute motion of the optical centre of the camera can be rewritten as follow with the utilization of unit orthogonality of rotation matrix

$$\overrightarrow{\mathbf{CC'}} = -R_{cw}^{\mathrm{T}}R_{c'c}^{\mathrm{T}}\mathbf{t}_{c'c} \tag{5}$$

Note that the $t_{c'c}$ term is a vector corresponding to the C so the additional transform $-R_{cw}^{T}$ is introduced into. We denote $\overrightarrow{WO}$ as o and $\overrightarrow{WC}$ as c equations (2)(3)(4)(5) yields

$$\mathbf{t}_{c'c} = -R_{c'w}L(\mathbf{c}-\mathbf{o}) \tag{6}$$

where $L=(\cos\theta-1)I+\sin\theta[\mathbf{n}]_\times$ is extracted and we will discuss it next. Based on above derivations in

(1)(6), the dynamic exterior orientation can be presented by transformation matrix in a compact and iterative form as

$$T_{c'w} = T_{c'c} T_{cw} \tag{7}$$

And they have an unified form as follow except with corresponding subscripts attached to them.

$$T = \begin{bmatrix} R & \mathrm{t} \\ 0^{\mathrm{T}} & 1 \end{bmatrix}.$$

It should be made clear that we update the component of translation directly to avoid complicated calculations, started from the second equation in (2) that

$$t_{c'w} = -R_{c'w}[L(c-o)+c] \tag{8}$$

According to our assumptions of the model, there are two inherent constraints between the parameters that

$$\begin{aligned} \|n\|_2^2 &= 1 \\ n \cdot (c-o) &= 0. \end{aligned} \tag{9}$$

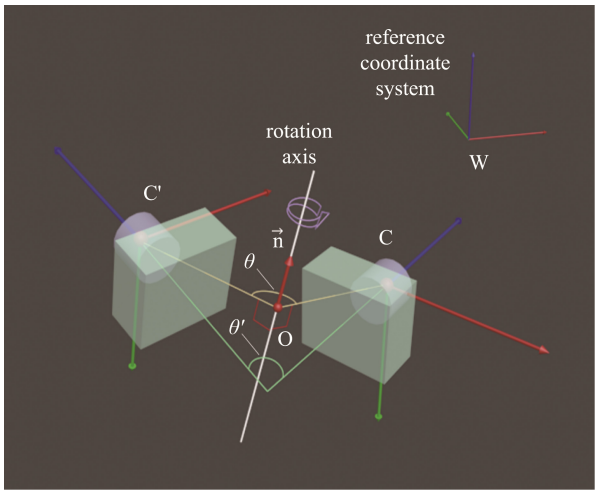


Fig.1 RVS Non-aligned Axis Deviation Model

3 Estimation Algorithms

To determine the rotational axis deviation parameters of the RVS model constructed in the previous section, the estimation method should be designed based on the specific system configuration of the actual application. For example, an RGB-D camera that provides depth measurement can simplify the estimation process to some extent. However, from a more general perspective, this paper proposes different types of calibration methods for general camera hardware configurations and conducts corresponding experiments.

3.1 Estimation with Stereo Camera Calibration

The hardware system setup for the experiment is shown in Fig.2. The stationary camera is fixed on a tripod and kept still during the calibration process. The rotating camera is fixed on a two-dimensional electric control gimbal and rotates sequentially with the gimbal around a certain axis. Assuming that the gimbal is rotated n groups of equal intervals relative to a certain rotation axis, and the angle interval between the rotations $\delta\theta$ is not less than 2°, the combination of the number of groups n and the equal angle interval $\delta\theta$ should ensure that the binocular stereo vision system always has sufficient common visual field (not less than 1/9 of the stationary camera's field of view). In fact, according to equations (1) and (9), the correspondence between relative rotation and axis-angle can be simplified into the following linear system of equations.

$$2\sin\boldsymbol{\theta} \cdot \boldsymbol{n} = \begin{bmatrix} \boldsymbol{r}_{32} - \boldsymbol{r}_{23} \\ \boldsymbol{r}_{13} - \boldsymbol{r}_{31} \\ \boldsymbol{r}_{21} - \boldsymbol{r}_{12} \end{bmatrix} \tag{10}$$

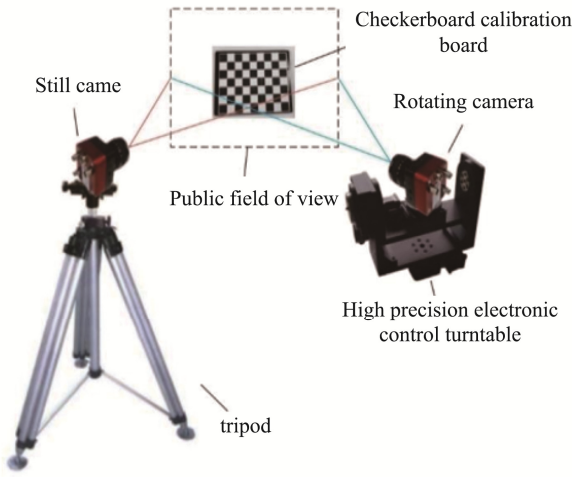


Fig.2 Schematic Diagram of RVS Calibration Based on Stereo Extrinsic Parameters

Similarly, if the coordinates of the camera's optical center before and after rotation are denoted as $\boldsymbol{c}(X,$

Y, Z) and $\boldsymbol{c}'(X', Y', Z')$ respectively, combined with the above equations (2) and (8) and after a simple transformation, we can obtain

$$\begin{bmatrix} \cos\theta-1 & -\sin\theta\cdot\boldsymbol{n}_z & \sin\theta\cdot\boldsymbol{n}_y \\ \sin\theta\cdot\boldsymbol{n}_z & \cos\theta-1 & -\sin\theta\cdot\boldsymbol{n}_x \\ -\sin\theta\cdot\boldsymbol{n}_y & \sin\theta\cdot\boldsymbol{n}_x & \cos\theta-1 \end{bmatrix}\boldsymbol{o}=\boldsymbol{s} \quad (11)$$

where A is the intermediate quantity introduced for the sake of a concise formal expression of the equation.

$$\begin{aligned} s &\equiv (L+I)c-c' \\ &= \begin{bmatrix} \cos\theta\cdot X-\sin\theta\cdot\boldsymbol{n}_z\cdot Y+\sin\theta\cdot\boldsymbol{n}_y\cdot Z-Y' \\ \sin\theta\cdot\boldsymbol{n}_z\cdot X+\cos\theta\cdot Y-\sin\theta\cdot\boldsymbol{n}_x\cdot Z-Y' \\ -\sin\theta\cdot\boldsymbol{n}_y\cdot X+\sin\theta\cdot\boldsymbol{n}_x\cdot Y+\cos\theta\cdot Z-Z' \end{bmatrix} \end{aligned} \quad (12)$$

The external parameters of the binocular camera changes are calibrated using the binocular stereo vision system during the intervals of the said equally spaced rotations (with the stationary camera as reference) to obtain the relative poses of the cameras before and after rotation.

$$\boldsymbol{R}_{c^ic^0}=\boldsymbol{R}_{c^ic^l}\boldsymbol{R}_{c^lc^0}=\boldsymbol{R}_{c^ic^l}\boldsymbol{R}^T_{c^0c^l} \quad (13)$$

The subscript $\boldsymbol{c}^0$ represents the initial position of the camera, while $\boldsymbol{c}^l$ corresponds to the stationary reference camera (left camera), and $\boldsymbol{c}^i(\boldsymbol{i}=1, 2, \ldots, \boldsymbol{n})$ represents the corresponding position of the camera after its group $\boldsymbol{i}$ rotation, with their respective optical center coordinates being $\boldsymbol{c}^0(X^0, Y^0, Z^0)$, $\boldsymbol{c}^l(X^l, Y^l, Z^l)$, and $\boldsymbol{c}^i(X^i, Y^i, Z^i)$. The overdetermined linear equation system is constructed incrementally in an additive manner as shown below.

$$\begin{aligned} &\boldsymbol{A}_{\mathrm{n}}\boldsymbol{n}=\boldsymbol{b}_n \\ &\boldsymbol{A}_{\mathrm{n}}=\begin{bmatrix} 2\sin\delta\theta & 0 & 0 \\ 0 & 2\sin\delta\theta & 0 \\ 0 & 0 & 2\sin\delta\theta \\ \vdots & \vdots & \vdots \\ 2\sin(\boldsymbol{n}\cdot\delta\theta) & 0 & 0 \\ 0 & 2\sin(\boldsymbol{n}\cdot\delta\theta) & 0 \\ 0 & 0 & 2\sin(\boldsymbol{n}\cdot\delta\theta) \end{bmatrix},\quad \boldsymbol{b}_n=\begin{bmatrix} \boldsymbol{r}^1_{32}-\boldsymbol{r}^1_{23} \\ \boldsymbol{r}^1_{13}-\boldsymbol{r}^1_{13} \\ \boldsymbol{r}^1_{21}-\boldsymbol{r}^1_{12} \\ \vdots \\ \boldsymbol{r}^n_{32}-\boldsymbol{r}^n_{23} \\ \boldsymbol{r}^n_{13}-\boldsymbol{r}^n_{31} \\ \boldsymbol{r}^n_{21}-\boldsymbol{r}^n_{12} \end{bmatrix} \end{aligned} \quad (14)$$

Find the least squares estimate of the direction vector by solving this system of linear equations.

$$\hat{\boldsymbol{n}}=(\boldsymbol{A}_n^T\boldsymbol{A}_n)^{-1}\boldsymbol{A}_n^T\boldsymbol{b}_n \quad (15)$$

Similarly, the obtained rotation direction vector estimates are subsequently substituted into the following set of equations to obtain the least-squares estimate of the vertical center The only difference is that the constraint described by the second equation in Eq.(9) is associated here

$$\begin{aligned} &\boldsymbol{A}_{\boldsymbol{O}}\boldsymbol{O}=\boldsymbol{b}_o \\ &\boldsymbol{A}_{\boldsymbol{O}}=\begin{bmatrix} \cos\delta\theta-1 & -\sin\delta\theta\cdot\boldsymbol{n}_z & \sin\delta\theta\cdot\boldsymbol{n}_y \\ \sin\delta\theta\cdot\boldsymbol{n}_z & \cos\delta\theta-1 & -\sin\delta\theta\cdot\boldsymbol{n}_x \\ -\sin\delta\theta\cdot\boldsymbol{n}_y & \sin\delta\theta\cdot n_x & \cos\delta\theta-1 \\ \vdots & \vdots & \vdots \\ \cos(\boldsymbol{n}\cdot\delta\theta)-1 & -\sin(\boldsymbol{n}\cdot\delta\theta)\cdot\boldsymbol{n}_z & \sin(\boldsymbol{n}\cdot\delta\theta)\cdot\boldsymbol{n}_y \\ \sin(\boldsymbol{n}\cdot\delta\theta)\cdot\boldsymbol{n}_z & \cos(\boldsymbol{n}\cdot\delta\theta)-1 & -\sin(\boldsymbol{n}\cdot\delta\theta)\cdot\boldsymbol{n}_x \\ -\sin(\boldsymbol{n}\cdot\delta\theta)\cdot\boldsymbol{n}_y & \sin(\boldsymbol{n}\cdot\delta\theta)\cdot\boldsymbol{n}_x & \cos(\boldsymbol{n}\cdot\delta\theta)-1 \\ \rho\cdot\boldsymbol{n}_x & \rho\cdot\boldsymbol{n}_y & \rho\cdot\boldsymbol{n}_z \end{bmatrix},\boldsymbol{b}_o=\begin{bmatrix} \boldsymbol{S}^1_x \\ \boldsymbol{S}^1_y \\ \boldsymbol{S}^1_z \\ \vdots \\ \boldsymbol{S}^n_x \\ \boldsymbol{S}^n_x \\ \boldsymbol{S}^n_x \\ \rho(\boldsymbol{n}\cdot\boldsymbol{c}^0) \end{bmatrix} \end{aligned} \quad (16)$$

Where ρ is the weight of the constraint equation, it should be set as large as possible (set to 10^5 in the experiment).

3.2 Estimation with Angel Measurement

The camera is initialized to capture satisfying images, and then holds the imaging settings constant. And we assume the intrinsic parameters to be invariable when the camera rotates. Thus, the estimation process can be separated where the change in pixels can be attributed to camera motion only. The intrinsic parameters composing into a matrix $\boldsymbol{K}$ are

calibrated by Zhang's method [18] with a checkerboard pattern. The pattern is subsequently used to estimate our rotation parameters with the help of OpenCV functions [19] to get the pixel coordinates of the corner points followed by distortion elimination. We make use of the reprojection error again this time for constructing the optimization problem of camera motion. A corner point of the checkboard pattern is denoted as $\boldsymbol{P}[\boldsymbol{X},\ \boldsymbol{Y},\ 0]^T$ when we define the world coordinate system being located at it.

We replace $\boldsymbol{P}$ with $\boldsymbol{Q}[\boldsymbol{X},\ \boldsymbol{Y},\ 1]^T$ and $\boldsymbol{R}_{c'w}$, $\boldsymbol{t}_{c'w}$ with $\boldsymbol{G}[\boldsymbol{r}_1,\ \boldsymbol{r}_2,\ \boldsymbol{t}]$ where $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ are the first two columns of the rotation matrix $\boldsymbol{R}_{c'w}$ and $\boldsymbol{t}$ is the translation vector $\boldsymbol{t}_{c'w}$. The initial EO of the camera can be solved as a Perspective-n-Points (PnP) problem at any proper location or just the last location of the camera calibration process. Then the simplified transformation matrix G can be derived from the model described in Section . Thus the estimated projection of $\boldsymbol{P}$ can be expressed forward as

$$\hat{\boldsymbol{q}} \sim \boldsymbol{KGQ} \tag{17}$$

Here the ~ sign represents it's an equation in the homogeneous sense for the projection point on the image plane is of 2 dimensions. And the DOF of variables is 4 for the two parametric vectors are both of 3 dimensions but with 2 equation constraints discussed above in (9). That means the minimum configuration solution includes 2 corresponding points with only 1 rotation motion. Nevertheless, we take the same setup strategy naturally as in bundle adjustment (BA) optimization process to mitigate the possible coupling of the above two parametric vectors. Assume we are given n images of the complete checkboard pattern at n discrete locations from camera rotations and the total number of corner points is m , we can obtain the following object function under maximum-likelihood estimate:

$$\min_{\boldsymbol{n},\boldsymbol{o}} \sum_{i=1}^{\boldsymbol{m}} \sum_{j=1}^{\boldsymbol{n}} \left\| \boldsymbol{q}^{ij} - \hat{\boldsymbol{q}}(\boldsymbol{n},\ \boldsymbol{o} \mid \boldsymbol{\theta}_i,\ \boldsymbol{K},\ \boldsymbol{Q}_j,\ \boldsymbol{R}_{cw},\ \boldsymbol{t}_{cw}) \right\|_2^2 \tag{18}$$

where $\hat{\boldsymbol{q}}\,(\boldsymbol{n},\ \boldsymbol{o}|\boldsymbol{\theta}_i,\ \boldsymbol{K},\ \boldsymbol{Q}_j,\ \boldsymbol{R}_{cw},\ \boldsymbol{t}_{cw})$ is the estimated projection of point $\boldsymbol{P}_j$ in image i corresponding to a rotation angle of θi according to (17) and $\boldsymbol{q}_{ij}$ is the actual observation. Combined with (9), the estimation task turns out to be solving a constrained nonlinear least square optimization problem. The initial value of $\boldsymbol{n}$ is simply set as $[1\ 0\ 0]^T$ and $[0\ 1\ 0]^T$ for tilt and pan rotation respectively, and o as c.

3.3 Estimation without Angle Measurements

It makes dynamic EO results unreliable just by the motion model of the camera formulated in Section when we lack of accurate angle measurements. Unless we are provided with useful observations which can establish the matches among rotations. But for the estimation of rotation parameters, we can utilize the particular motion paradigm that the trajectory of the camera is part of a circle in space. We let $\{\boldsymbol{R}_{\mathrm{i}},\ \boldsymbol{t}_i|\boldsymbol{i}=0,\ 1,\ 2,\ \ldots,\ \boldsymbol{n}\}$ be the solutions through PnP and the $\boldsymbol{R}_0$, $\boldsymbol{t}_0$ denotes the extrinsic parameters of the camera before it rotates. Then coordinates $\{\boldsymbol{c}_i\}$ of the optical centers $\{\boldsymbol{C}_i\}$ of the camera are obtained according to (2). The circle in space can be mathematically represented by

$$\begin{cases} \left\| \boldsymbol{p} - \boldsymbol{o} \right\|_2^2 = \boldsymbol{e}^2 \\ \boldsymbol{n} \cdot (\boldsymbol{p} - \boldsymbol{o}) = 0 \end{cases} \tag{19}$$

where $\boldsymbol{p}$ denotes the coordinate of an arbitrary point on the circle and e is the radius of the circle which reflects the eccentric distance of a RVS. As illustrated in Fig.3, the practical points $\{\boldsymbol{c}_i\}$ are of little probability locating exactly at the circle due to noise from measurements and the PnP solutions. We use $\{\overline{\boldsymbol{c}}_i\}$ to denote the coordinates of points $\{\overline{\boldsymbol{C}}_i\}$ which satisfy the above circle equation nearest to the practical points, and $\{\boldsymbol{d}_i\}$ to denote those of the perpendicular intersections $\{\boldsymbol{D}_i\}$ which are formed from points $\{\boldsymbol{c}_i\}$ to the hypothetic plane formulated by the second equation in (19). Then similarly, the problem of circle fitting transforms to minimizing the Euclidean distance of corresponding points and thus the objective function here is

$$\min \sum_{i=0}^{\boldsymbol{n}} \left\| \overrightarrow{\boldsymbol{C}_i \overline{\boldsymbol{C}}_i} \right\|_2^2 \tag{20}$$

It's a basic equation that

$$\left\| \overrightarrow{\boldsymbol{C}_i \overline{\boldsymbol{C}}_i} \right\|_2^2 = \left\| \overrightarrow{\boldsymbol{C}_i \boldsymbol{D}_i} \right\|_2^2 + \left\| \overrightarrow{\overline{\boldsymbol{C}}_i \boldsymbol{D}_i} \right\|_2^2 \tag{21}$$

According to algebraic geometry it's also quite sim-

ple that

$$\left\|\overrightarrow{C_iD_i}\right\|_2=\frac{|\boldsymbol{v}_{2i}|}{\|\boldsymbol{n}\|_2}=|\boldsymbol{v}_{2i}| \tag{22}$$

and

$$\begin{aligned}\left\|\overrightarrow{\bar{C}_iD_i}\right\|_2&=\left|\left\|\overrightarrow{OD_i}\right\|_2-\left\|\overrightarrow{O\bar{C}_i}\right\|_2\right|\\&=\left|\sqrt{e^2+\boldsymbol{v}_{1i}-\boldsymbol{v}_{2i}^2}-e\right|\end{aligned} \tag{23}$$

where $\boldsymbol{v}_{1i}$ and $\boldsymbol{v}_{2i}$ are residuals when $\boldsymbol{c}_i$ of point $\boldsymbol{C}_i$ is substituted into the first and second equation in (19) respectively. In this case where the sampled points are sparse and concentrated in a small part of a virtual circle, we impose the following constraint on the radius to avoid the predictable overparameterization.

$$\boldsymbol{e}=\frac{1}{\boldsymbol{n}+1}\sum_{i=0}^{\boldsymbol{n}}\|\boldsymbol{c}_i-\boldsymbol{o}\|_2 \tag{24}$$

Then in the same way, the problem is also a constrained least-square optimization one with simultaneous equations listed in this subsection and the constraints described above except with the change from the second equation in (9) to

$$\boldsymbol{n}\cdot(\frac{1}{\boldsymbol{n}+1}\sum_{i=0}^{\boldsymbol{n}}\boldsymbol{c}_i-\boldsymbol{o})=0 \tag{25}$$

The geometrical interpretation of (24) and (25) are somehow reasonable situations where the sampled points are symmetrical distributed around the sphere and plane respectively, which are defined by the two equations in (19).

It needs to be clarified that the parametric vectors $\boldsymbol{n}$ and $\boldsymbol{o}$ are estimated separatly here to reduce the impact of uncertain coordination errors of points $\{\boldsymbol{C}_i\}$. We take advantage of the convenient transformation from a rotation matrix to a rotation vector. Rotation matrix sets $\{\boldsymbol{R}_i|\boldsymbol{i}=0,1,2,\ldots,\boldsymbol{n}\}$ are converted to $\{\boldsymbol{R}_j\boldsymbol{R}_i^T\,|\,\boldsymbol{i},\boldsymbol{j}=0,1,2,\cdots,\boldsymbol{n};\boldsymbol{i}<\boldsymbol{j}\}$ and then the corresponding rotation vectors $\{\boldsymbol{r}_{ij}\}$. For the simplicity of representation, we denote the vector normalization operation as N and the direction of rotation axis can be extracted by

$$n=N(\sum_{i,j}\rho_{ij}N(r_{ij})) \tag{26}$$

where ρ_{ij} are weights attached to different image index i, j. We make it simply be $\rho_{ij}=(j-i)^2$ to put more trust on image pairs of larger angle interval.

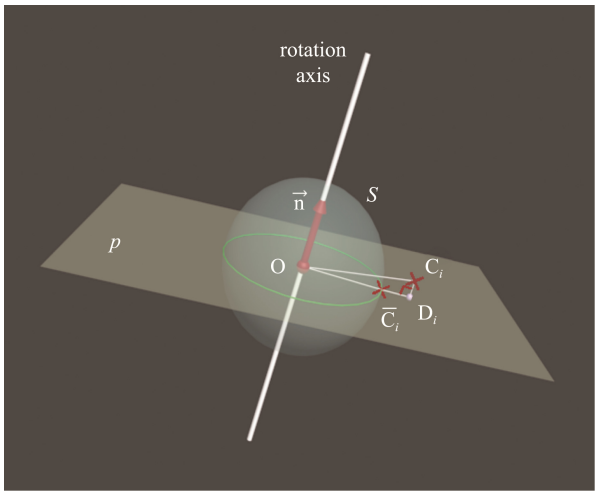


Fig.3 Schematic Diagram of Spatial Circle Fitting for RVS Calibration

4 Experiments

4.1 Based on Stereo Extrinsic Calibration

The corresponding experiments were carried out according to the described method, setting the equal angle interval of the PT rotation in both pan(Y) and tilt(X) directions to 5°, and obtaining 4 and 5 sets of episodic calibrations for binocular stereo vision co-viewing cases, respectively. Then, we substitute The results of the least-squares estimation of the parameters obtained by solving the linear equation system are shown in Table 1 below.

The calibration results of the pan axis (Y-axis) in the table exhibit significant errors (highlighted in red), which are due to numerical issues arising in the direct solution of the linear equation system for the camera-gimbal system when the corresponding axis heights in the world coordinate system coincide. As described herein, this method using equally spaced angle intervals set to 5° only allows for 4 and 5 groups on the two axes, respectively, which is insufficiently redundant for constructing an overdetermined linear equation system. This raises potential contradictions in the axis estimation process. When a sufficient number of groups is used, the angle interval is restricted to a

Table 1 Estimation Results Based on Stereo Extrinsic Calibration

		Tilt		Pan	
		PT	Camera	PT	Camera
rotation vector	n_x	0.785864	0.811708	–0.000993	0.0074508
	n_y	0.11122	0.11599	1	0.99201
	N_z	0.60832	0.57243	–2.95e–15	–0.125936
Vertical center	O_x	1325.21	1270.05	6762.63	600.466
	O_y	249.736	76.2511	78.0451	71.9265
	O_z	–469.741	–366.757	4219.95	–672.546

certain range, which results in a high risk of "ill-conditioning" of the previously extracted matrix L, particularly when the relevant parameters n and o differ by more than 3-4 orders of magnitude. This risk leads to more severe numerical issues.

4.2 Based on Minimize Reprojection Error

We tested the proposed model and estimation methods on a RVS composed of an electrical turntable and a camera with a resolution of 2000 × 1000. The results of the camera calibration internal obtained with the MATLAB Toolbox [20] are shown in Table 2 and the overall estimation in the form of reprojection error to obtain the evaluation results is shown in Fig.4. Fig.5 shows the captured real image of major parts of the experimental setup.

Table 2 Intrinsic Outputs of Camera Calibration

Intrinsic Matrix	$\boldsymbol{K}=\begin{bmatrix}2232.3188 & 0 & 0\\ 0 & 2233.0692 & 0\\ 1051.3174 & 518.7745 & 1\end{bmatrix}$
Distortion Factor	$k_1=-0.107528784$, $k_2=0.23666415$ $p_1=-0.00114327$, $p_2=-0.0004598$

Sampled images, as illustrated in Fig.6, were generated by tilt (axis X) and pan (axis Y) motion of turntable while the reused calibration pattern was kept absolutely stationary. The 150 inner corner-points (10 rows × 15 columns) are guaranteed to be in the FOV of the camera throughout the calibration process. In order to maintain consistency in the following discussion, image sequences are indexed respectively from 0 to 8 and 0 to 13 for pan and tilt, with the index 0 corresponding to the referenced frame. And units of data in $\boldsymbol{o}$ and $\boldsymbol{t}$ are both mm. The rotation angles were controlled to be uniformly 2° among adjacent frames for both pan and tilt (it can be adjusted flexibly according to the angle of view).

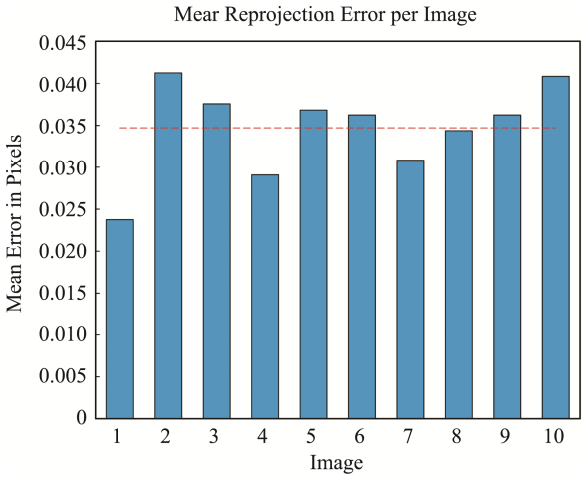


Fig.4 Reprojection Error of Camera Calibration

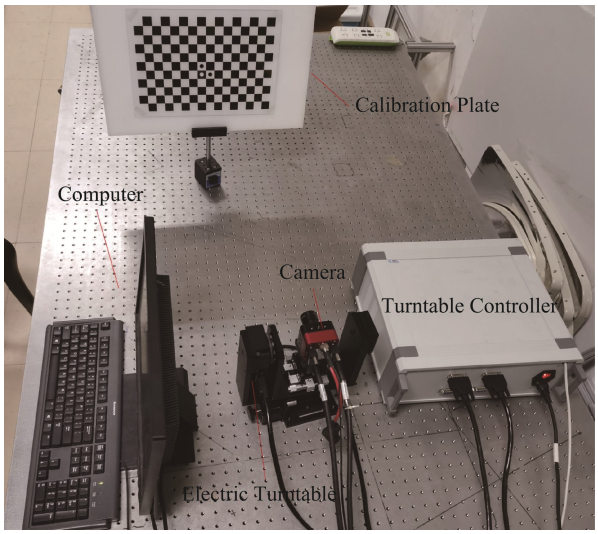
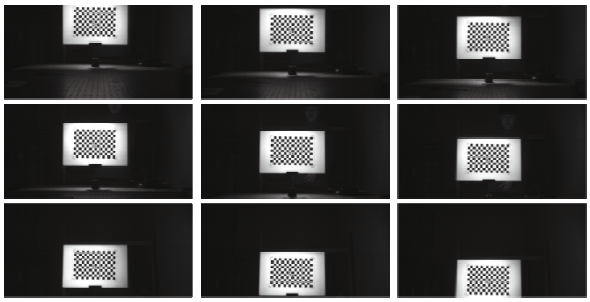
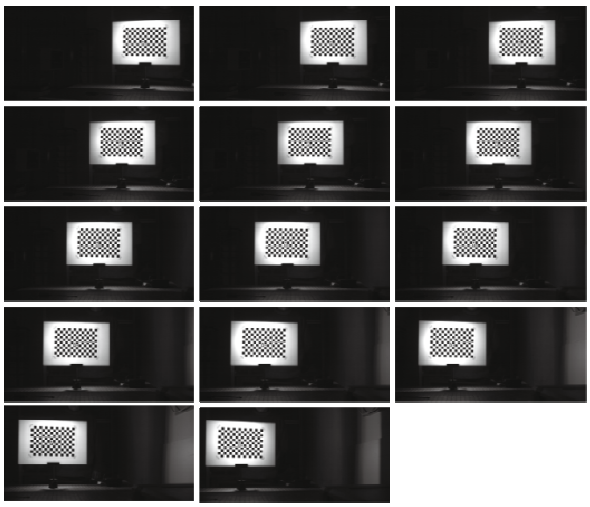


Fig.5 Schematic Diagram of Spatial Circle Fitting for RVS Calibration



(a) Tilt Axis



(b) Pan Axis

Fig.6 PT Turntable Rotates the Camera Step by Step to Capture Images

The final estimation results are listed in Table 3. For the ultimate goal of axis estimation is to extend the sensing range of the camera, we took the same evaluation criteria as those in the camera calibration and then made straightforward comparisons. The estimation results of system parameters were substituted into the model formulated in Section to reconstruct the EOs of the camera. It was performed by taking the EO of the first frame of the image sequence as referenced prior.

Table 3 Estimation Results Based on Minimizing Reprojection Error

	n	o
Tilt (X axis)	$\begin{bmatrix}-0.99993\\-0.00274\\-0.01140\end{bmatrix}$	$\begin{bmatrix}60.3397\\-55.3728\\-1633.9647\end{bmatrix}$
Pan (Y axis)	$\begin{bmatrix}0.00128\\-0.99972\\0.02347\end{bmatrix}$	$\begin{bmatrix}64.8049\\-53.7632\\-1640.4430\end{bmatrix}$

Through the difference with the results of PnP, the reconstruction outputs were verified as shown in Table 4.

Furthermore, combining the results of camera intrinsic parameters and the reconstructed EO, the reprojections of the pattern were obtained. We calculated the mean reprojection error of each frame, and compared it of the referenced frame with the holistic mean reprojection error of those exterior reconstructed, as illustrated in Fig.7.

Similar to the output evaluation of the extrinsic parameters in the camera calibration in MATLAB, the ExtrinsicsErrors module in estimationErrors object, our reconstructed EOs according to (7) are compared to the PnP solutions directly. The results listed in Table 4 turn out to be of the same magnitude as those in the ExtrinsicsErrors. Mean reprojection errors of pan and tilt from dynamic reconstructed EO are 0.14 and 0.08 pixels in Fig.6, with it of the camera calibration result is about 0.03 pixels. It is intuitive that the more of the absolute rotation angle (which means closer to the edge of the image), the greater of the mean reprojection error. It may be caused by the insufficient distortion calibration. And we can perform outliers removal of some "poor" frames (just like the frames corresponding to absolute angles of the turntable of 9, 11, 13). Or replace the referenced frame with those that more orthogonal to the calibration plate.

4.3 Based on Spatial Circle Fitting

In the image acquisition experiments described earlier, the spatial circle fitting algorithm was employed to estimate the rotation axis deviations when the Pan -Tilt readings were unknown. The parameters obtained are shown in Table 5 and were compared with those in Table 3 by analytical means, since the EO reconstruction process could not be performed due to the absence of angular measurements.

The reprojection error-based method was chosen as a benchmark for quantitative evaluation. The results showed that the circle fitting estimation method yielded comparable results to the reprojection error-based method when angular measurements were available (relative error around 2%). However, it was found to be more sensitive to noise. By increasing the

number of sample points (e.g. by reducing the rotation angle interval) and using outlier detection techniques like RANSAC, the estimation results of this method can become more stable.

Table 4 Deviation between RVS Dynamic EO Reconstruction Results and PnP Solutions

Serial Number	Tilt (X)					
	δr			δt		
	δr_x	δr_y	δr_z	δt_x	δr_x	δr_x
1	–2.278e–3	7.81e–4	9.1e–5	–0.0327	–0.0804	–0.5882
2	1.781e–3	1.375e–3	2.2e–5	–0.0351	–0.0924	–0.2685
3	2.019e–3	–6.50e–4	–4.3e–5	0.0511	–0.1704	—0.0474
4	–1.44e–3	2.494e–3	–7.2e–5	0.0042	–0.0150	—0.7796
5	–1.991e–3	1.263e–3	–9.71e–4	0.0525	–0.0262	–0.5882
6	–6.73e–4	3.238e–3	–3.98e–4	–0.0654	–0.0864	–1.4897
7	–1.409e–3	1.401e–3	–2.27e–4	–0.0304	–0.2170	–1.2415
8	7.0e–5	2.955e–3	–5.40e–4	–0.0295	–0.2031	–1.7966
Serial Number	Pan (Y)					
	δr^T			δt^T		
	δr_x	δr_y	δr_z	δt_x	δr_x	δr_x
1	1.589e–3	–4.13e–4	–2.96e–4	0.3832	–0.1447	0.4733
2	–4.009e–3	9.23e–4	6.77e–4	0.0126	–0.0929	–0.4973
3	–4.1e–5	2.754e–3	1.63e–4	0.3335	–0.2602	0.2174
4	2.055e–3	–2.47e–4	–1.52e–4	0.2150	–0.1811	0.8470
5	–4.84e–4	1.309e–3	6.2e–5	0.4606	–0.3395	0.1905
6	3.418e–3	1.408e–3	2.51e–4	0.6882	–0.2072	1.3807
7	3.057e–3	–3.0e–5	1.42e–4	0.8491	–0.2257	1.4190
8	–1.30e–4	7.13e–4	2.63e–4	0.5812	–0.1215	1.2074
9	–8e–6	1.139e–3	3.53e–4	0.1877	–0.1805	0.8890
10	1.682e–3	1.60e–4	6.82e–4	–0.3065	–0.0654	1.3625
11	–2.69e–4	–1.168e–3	3.74e–4	–0.2331	–0.1446	1.4577
12	–1.411e–3	–1.369e–3	2.67e–4	–0.5972	0.2962	1.0826
13	–3.276e–3	–1.45e–4	–1.4e–5	–2.1034	0.0991	0.1326

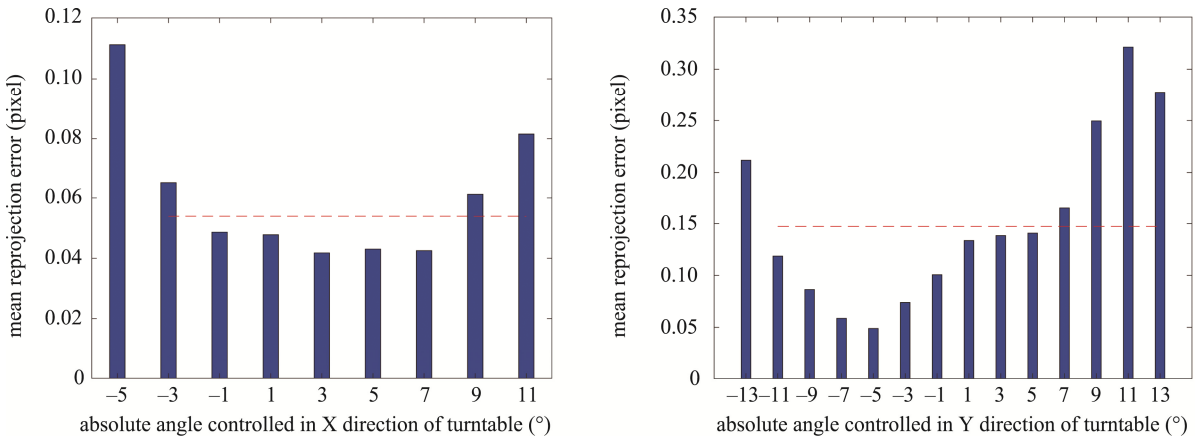


Fig.7 Reprojection Error of RVS Calibration Results Evaluation

Table 5 Estimation Results Based on Spatial Circle Fitting

	n	o
Tilt (X axis)	$\begin{bmatrix}-0.99990\\-0.00950\\-0.01075\end{bmatrix}$	$\begin{bmatrix}61.9094\\-47.8969\\-1600.5692\end{bmatrix}$
Pan (Y axis)	$\begin{bmatrix}0.00358\\-0.99974\\0.02259\end{bmatrix}$	$\begin{bmatrix}62.2023\\-53.9735\\-1643.9392\end{bmatrix}$

5 Conclusion

We proposed a generic space-equivalent non-aligned axis-centre model of RVS for dynamic EO of the camera, and presented two types of estimation methods of the rotation axis specified to different hardware conditions based on the same model. And the calibration experiments are conducted on a relatively common pan & tilt (PT) system. It is designed to be integrated with the camera calibration procedure reflected in three aspects. First, the calibration pattern is shared with. Then the complete imaging system can be recovered with again only a few images of the pattern added. And by combining the whole calibration results of the system, we can take the same criteria with that of the camera calibration to evaluate the output. Benifiting from these, the estimation methods are shown to be accurate and efficient.

It is indicated that the referenced image selected has an impact on the estimation method of minimizing reprojection error from both theoretical model and experimental results. But this influence is not significant due to the robustness of the multi-view multi -point BA, similar to the mainstream setup of camera calibration. This is not the case for the estimation method of circle fitting, however, especially for such a narrow range of samples. In addition, the indeterminate combined errors of calculated coordinates of sampled points from PnP results make it even worse. A fairly straightforward way to minimize these negative effects is to establish a wider calibration target with encoding features to extend range of sampling points. We plan to investigate approaches and implement comparable experiments on applications such as rotary scanning 3D reconstruction and rotation servo control based on inverse kinematics, to evaluate and improve the model and estimation method

6 Acknowledgments

The authors appreciate the support of the National Natural Science Foundation of China (No.52175504 and51927811), the Fundamental Research Funds for the Central Universities of China (No.PA2022GDSK0074), and the National Key Research and Development Program of China (No.2022CSJGG1303).

References

[1] Haque, A.M. (2013)Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera. In: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. Krakow, Poland: IEEE, pp.443-448.

[2] Bimbo, A.D. (2010). Exploiting distinctive visual landmark maps in pan–tilt–zoom camera networks. Computer Vision and Image Understanding, 114(6), pp.611-623.

[3] Furukawa, T. (2017). Multi-stage Bayesian target estimation by UAV using fisheye lens camera and pan/tilt camera. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver, BC, Canada: IEEE, pp.4167–4172.

[4] Wang, Z Y. (2012). Cooperative Target Tracking Control of Multiple Robots. IEEE Transactions on Industrial Electronics, 59(8), pp. 3232-3240.

[5] Chen, H. (2014). A novel pan-tilt camera control approach for visual tracking. In: Proceeding of the 11th World Congress on Intelligent Control and Automation. Shenyang, China: IEEE, pp.2860-2865.

[6] Vadakkepat, P. (2006). Improved Particle Filter in Sensor Fusion for Tracking Randomly Moving Object. IEEE Transactions on Instrumentation and Measurement, 55(5), pp.1823-1832.

[7] Hayman, E. (2003). The effects of translational misalignment when self-calibrating rotating and zooming cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8), pp.1015-1020.

[8] Alvarez, L. (2012). Mathematical Models for the Cali-

bration of Cameras Mounted on a Tripod Using Primitive Tracking. In: Proceedings of the 9th International Conference on Image Analysis and Recognition. Aveiro, Portugal: Springer, pp.304-311.

[9] Byun, J. (2018). Accurate control of a pan-tilt system based on parameterization of rotational motion. In: Proceeding of the 39th European Association Computer Graphics Conference EUROGRAPHICS.

[10] Davis, J. (2003). Calibrating pan-tilt cameras in wide-area surveillance networks. In: Proceedings Ninth IEEE International Conference on Computer Vision. Nice, France: IEEE, pp.144-149.

[11] Wu, Z. (2012). Keeping a pan-tilt-zoom camera calibrated. IEEE transactions on pattern analysis and machine intelligence, 35(8), pp.1994-2007.

[12] Li, Y. (2015). Method for pan-tilt camera calibration using single control point. Journal of the Optical Society of America A, 32(1), pp.156-163.

[13] Nagayoshi, H. (2014). Estimating camera pose using trajectories generated by pan-tilt motion. In: 2014 2nd International Conference on 3D Vision. Tokyo, Japan: IEEE, pp.561-568.

[14] Chen, X. (2020). An Integrated Two-Pose Calibration Method for Estimating Head-Eye Parameters of a Robotic Bionic Eye. IEEE Transactions on Instrumentation and Measurement, 69(4), pp.1664-1672.

[15] Rebello, J. (2017). Autonomous active calibration of a dynamic camera cluster using next-best-view. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver, BC, Canada: IEEE, pp.1484-1489.

[16] Chen, P. (2014). Rotation axis calibration of a turntable using constrained global optimization. Optik, 125(17), pp.4831-4836.

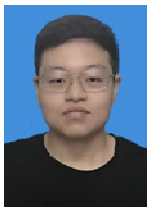
[17] Szeliski, R.: Computer Vision: Algorithms and Applications. Springer-Verlag, London, UK (2011).

[18] Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), pp.1330–1334.

[19] Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000).

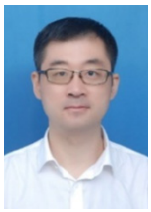
[20] Bouguet, J.Y.: Camera calibration toolbox for matlab. Www.vision.caltech.edu/bouguetj (2013).

Author Biographies



YAN Guangzong received his B.Sc. degree from Northeast Petroleum University in 2021. He is now a M.Sc. candidate in Hefei University of Technology. His main research interests include large-scale visual measurement and defect detection.

E-mail: 2021170012@mail.hfut.edu.cn



ZHANG Jin received his Ph.D. degree from Tianjin University in 2010. He is now a professor and Ph.D. supervisor in Hefei University of Technology. His main research interests include vision measurement, dynamic test, etc.

E-mail: zhangjin@hfut.edu.cn