

Depth Adaptive Zooming Visual Servoing for a Robot with a Zooming Camera

Regular Paper

Jing Xin^{1,*}, Kemin Chen¹, Lei Bai¹, Ding Liu¹ and Jian Zhang²¹ School of Automation and Information Engineering, Xi'an University of Technology, P.R. China² Faculty of Engineering and Information Technology, University of Technology (UTS), Sydney, Australia

* Corresponding author E-mail: xinj@xaut.edu.cn

Received 22 Mar 2012; Accepted 22 Oct 2012

DOI: 10.5772/54566

© 2013 Xin et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract To solve the view visibility problem and keep the observed object in the field of view (FOV) during the visual servoing, a depth adaptive zooming visual servoing strategy for a manipulator robot with a zooming camera is proposed. Firstly, a zoom control mechanism is introduced into the robot visual servoing system. It can dynamically adjust the camera's field of view to keep all the feature points on the object in the field of view of the camera and get high object local resolution at the end of visual servoing. Secondly, an invariant visual servoing method is employed to control the robot to the desired position under the changing intrinsic parameters of the camera. Finally, a nonlinear depth adaptive estimation scheme in the invariant space using Lyapunov stability theory is proposed to estimate adaptively the depth of the image features on the object. Three kinds of robot 4DOF visual positioning simulation experiments are conducted. The simulation experiment results show that the proposed approach has higher positioning precision.

Keywords Zooming Control, Invariant Visual Servoing, Depth Adaptive Estimation, Lyapunov Stability Theory

1. Introduction

The introduction of visual servoing techniques would increase the diversity and extend the application fields of a robot. The aim of robot visual servoing is to control the relative pose of the robot's end-effector, with respect to the manipulated object, using real-time visual information captured from a camera, which is either fixed in the robot work space (eye-on-hand configuration) or mounted at the robot's end-effector (eye-in-hand configuration). In the last few years, many robot visual servoing methods have been proposed to solve the problem of robot positioning and tracking with respect to the manipulated object in the dynamic and unknown environment. These methods can be roughly classified into three types according to the definition of the system error function: position-based visual servoing (PBVS), image-based visual servoing (IBVS), hybrid visual servoing (HVS or 2.5D VS). Each kind of visual servoing method has its own pros and cons. A comprehensive review paper about robot visual servoing can be found in [1].

The basic requirement of these visual servoing algorithms is to keep the observed object in the field of view of the camera during the visual servoing (also known as the visibility problem). When the robot initial position is far

from the desired position, the object could leave the camera field of view during the visual servoing. This will cause the visual servoing task to fail and limit its further application [2, 3]. There are two kinds of common solutions to this view visibility problem: trajectory planning and active vision techniques such as active zoom control. The main idea of trajectory planning is to plan the trajectories of a set of features points on the object in the image space and then to track these trajectories. At present, many trajectory planning methods have been proposed. In general these methods would be roughly classified into four types: image space-based planning [5], global path planning [6], optimization-based planning [7] and potential field-based planning [8]. Tracking a planned trajectory can keep the change of the robot pose in a small range. Thus, it would be possible to keep the object in the camera field of view by enforcing such constraints on the trajectories [4] and the visibility problem could be solved. However, this kind of method is sensitive to the change of the camera intrinsic parameters. A comprehensive review paper about trajectory planning can be found in [9].

Another solution is to combine the active vision technique with the robot visual servoing system, e.g., an active zoom control technique. The main idea of active zooming could be described as follows:

- *Zooming out* when one of the interest points, which are also called the features points on the object, is close to leaving the image.
- *Zooming in* when the interest points are well centred in the image, to get good resolution of the interesting points.

It can be seen that active zoom control could not only solve the view visibility problem, but also improve the accuracy of the features extraction from the interest object and the precision of the whole robot visual servoing system [2]. This has been widely used in the computer vision field [10, 11]. For example, Kumar et al. constructed a stereo vision system through the two pant-tilt-zoom (PTZ) cameras and localized a moving target precisely in a complex and large environment [11]. If we combine the active zooming idea with the robot visual servoing method, we could solve the view visibility problem in robot visual servoing and improve the accuracy of the whole visual servoing system. However, it is difficult to apply this active vision technique directly to traditional visual servoing approaches because traditional visual servoing approaches are based on a teaching-by-showing technique. This means that we must first store the desired image captured in the desired position and then control the robot, starting at another position than the desired position, where the current image coincides with the desired image. In other words, traditional visual servoing methods are “camera-dependent”. To solve this “camera-dependency” problem, one good solution is to employ

the invariant visual servoing (or intrinsic-free visual servoing) method proposed in [12]. The key idea of invariant visual servoing is to construct a system error function using a projective invariant property. The error function is invariant to the changes in camera intrinsic parameters and only relevant to the relative position and pose between camera and object. An image Jacobian matrix in the error function describes the relationship between the camera velocity and the image features error in the invariant space. This invariant visual servoing approach provides a good methodology for controlling the robot with a zooming camera, but it cannot be used on a planar object. In addition, it doesn't consider the focus length control problem. This means that we can control the robot using a different focus length to the one used in the learned stage, but we can't control or change the focus length during the visual servoing. In order to deal with this problem, a zooming visual servoing method is proposed in [13]. This method allows us to control the focus length during the visual servoing by using a focus length control strategy and then recover the focus length value corresponding to the desired image at the end of the visual servoing. In addition, this method can also solve the planarity problem from the method proposed in [12]. However, the methods proposed in [12 and 13] do not take into account depth estimation problems and estimate the object depth at the very beginning of the visual servoing, then keep it constant throughout. However, the image Jacobian matrix depends on the depth parameter, which is a time-varying parameter, so a constant depth estimation value could only locally guarantee the convergence of the method. In order to further improve the invariant visual servoing method proposed in [12 and 13], we proposed a nonlinear depth adaptive estimation algorithm using Lyapunov stability theory in the invariant space. Some depth estimation algorithms using Lyapunov stability theory have been proposed in the past years [14, 15]. These estimation methods can work well in the case of the fixed camera intrinsic parameters but can not be used in the case of the variant camera intrinsic parameters. In this paper we extend these depth estimation methods to zooming visual servoing, where the camera intrinsic parameter changes during the visual servoing and gives a detailed derivation. Finally, we design a zooming visual servoing strategy for a manipulator robot with a zooming camera. Simulation results show that our approach has a better convergence performance of the image error.

The remaining part of this paper is organized as follows. The basic principle of the invariant visual servoing method is briefly introduced in Section 2. The details of our algorithm are described in Section 3, where we propose an efficient depth adaptive zooming visual servoing approach. Simulation results and analysis are given in Section 4. Section 5 will conclude the paper by discussing the proposed approach and suggest some improvements for further research.

2. Basic Principles of Invariant Visual Servoing

In this section, we briefly introduce the basic principles of the invariant visual servoing method, related space notations and a pin-hole imaging model.

2.1 Space Notations

The 3D points, with homogeneous coordinates $\chi_i = (X_i, Y_i, Z_i, 1) (i = \{1, 2, \dots, n\})$ are projected in the absolute camera frame F^* to the points $\mathbf{m}_i^* (i \in \{1, 2, \dots, n\}) \in P^2$. These points are projected in the current camera frame F to the points $\mathbf{m}_i (i \in \{1, 2, \dots, n\}) \in P^2$. Then according to the projection model:

$$\mathbf{m}_i = \frac{1}{Z_i} [\mathbf{R}^* \ \mathbf{t}^*] \chi_i = (x_i, y_i, 1) \quad (1)$$

where Z_i is the positive depth and \mathbf{R}^* and \mathbf{t}^* are the rotation and translation, respectively, between F^* and F .

2.2 Camera Model

The image point $\mathbf{p} = [u \ v \ 1]^T$ is given by a pinhole camera in the current frame F . The relationship between \mathbf{p} and \mathbf{m} can be written as:

$$\mathbf{p}(\mathbf{Z}, \mathbf{K}) = \mathbf{K}\mathbf{m}(\mathbf{Z}) \quad \mathbf{p} \in I(\mathbf{Z}, \mathbf{K}) \quad (2)$$

where \mathbf{K} is the camera intrinsic parameter matrix:

$$\mathbf{K} = \begin{bmatrix} f k_u & -f k_u \cot(\theta) & u_0 \\ 0 & f k_v / \sin(\theta) & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

f is the focal length (in metres), k_u and k_v separately are the magnifications in the \bar{u} and \bar{v} direction (in pixels/m), u_0 and v_0 are the coordinates of the principle point (in pixels) and θ is the angle between \bar{u} and \bar{v} axes.

2.3 Basic Principle of the Invariant Visual Servoing

The main idea of the invariant visual servoing proposed in [12] is to construct an error function, which is independent of the changes in camera intrinsic parameters and θ is only dependent on the relative pose of the camera/robot, with respect to an observed object on the invariant space. The image Jacobian matrix in the error function describes the relationship between the camera velocity and image features error in the invariant space. Then the visual servoing law can be designed by the error function. Therefore, the construction of the invariant space and image Jacobian matrix are two important elements for invariant visual servoing.

2.3.1 Construction of the Invariant Space

The aim of the construction of the invariant space is to keep the features points invariant to the camera intrinsic

parameters. The whole construction process can be divided into five steps as follows.

Step 1: to obtain the pixel coordinates of the target in the desired frame.

Step 2: to chose three non-collinear 3D points χ_1, χ_2, χ_3 among the n non-coplanar points $\chi_i (i \in \{1, 2, \dots, n\})$, which are visible on the observed object.

Step 3: to obtain the corresponding three normalized projective points $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ (unit:m) in the current frame and another three points $\mathbf{m}_1^*, \mathbf{m}_2^*, \mathbf{m}_3^*$ (unit:m) in the desired frame respectively.

Step 4: to compute the corresponding image pixel coordinate $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ and $\mathbf{p}_1^*, \mathbf{p}_2^*, \mathbf{p}_3^*$ respectively according to the camera model defined in (2), as follows:

$$\mathbf{Q}(\mathbf{Z}, \mathbf{K}) = \mathbf{K}\mathbf{M}(\mathbf{Z}) \quad (4)$$

$$\mathbf{Q}^*(\mathbf{Z}^*, \mathbf{K}^*) = \mathbf{K}^* \mathbf{M}^*(\mathbf{Z}^*) \quad (5)$$

where $\mathbf{M} = [\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3]$, $\mathbf{M}^* = [\mathbf{m}_1^* \ \mathbf{m}_2^* \ \mathbf{m}_3^*]$
 $\mathbf{Q} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \mathbf{p}_3]$, $\mathbf{Q}^* = [\mathbf{p}_1^* \ \mathbf{p}_2^* \ \mathbf{p}_3^*]$.

Non-singular matrices \mathbf{Q} and \mathbf{Q}^* are called as invariant spaces transformation matrices.

Image feature point \mathbf{p} and \mathbf{p}^* in the original image space can be converted into the invariant image feature point and \mathbf{q}^* in the invariant space via invariant spaces transformation matrix \mathbf{Q} and \mathbf{Q}^* as follow

$$\mathbf{q}(\mathbf{Z}) = \mathbf{Q}^{-1} \mathbf{p} = \mathbf{M}^{-1} \mathbf{K}^{-1} \mathbf{K} \mathbf{m} = \mathbf{M}^{-1} \mathbf{m} \quad (6)$$

$$\mathbf{q}^*(\mathbf{Z}^*) = \mathbf{Q}^{*-1} \mathbf{p}^* = \mathbf{M}^{*-1} \mathbf{K}^{*-1} \mathbf{K}^* \mathbf{m}^* = \mathbf{M}^{*-1} \mathbf{m}^* \quad (7)$$

From the equations above we know that both \mathbf{q} and \mathbf{q}^* do not depend on the camera intrinsic parameters, but are only related to the camera position with respect to the observed object and the 3D structure itself. Now we can use the error in invariant space to control the camera's motion, even if the camera intrinsic parameters change during the visual servoing.

2.3.2 Construction of the Image Jacobian Matrix in the Invariant Space

Just like in the IBVS (image-based visual servoing) approach, the camera is now controlled in the invariant space. The derivative of the vector \mathbf{q} is

$$\dot{\mathbf{q}} = \mathbf{J}_q \mathbf{v}_c \quad (8)$$

\mathbf{J}_q is the (3*6) Jacobian matrix in the invariant space and is similar to the image Jacobian matrix. After the derivative of equation (6) we get:

$$\dot{\mathbf{q}}_i = \frac{d\mathbf{Q}^{-1}}{dt} \mathbf{p}_i + \mathbf{Q}^{-1} \dot{\mathbf{p}}_i$$

By plugging equation (6) into the above equation, we can get:

$$\dot{\mathbf{q}}_i = \frac{d\mathbf{Q}^{-1}}{dt} \mathbf{p}_i + \mathbf{Q}^{-1} \dot{\mathbf{p}}_i = \frac{d\mathbf{Q}^{-1}}{dt} \mathbf{Q} \mathbf{q}_i + \mathbf{Q}^{-1} \dot{\mathbf{p}}_i \quad (9)$$

Since $\frac{d\mathbf{Q}^{-1}}{dt} \mathbf{Q} = -\mathbf{Q}^{-1} \frac{d\mathbf{Q}}{dt}$, by plugging this into equation (9) we could obtain:

$$\begin{aligned} \dot{\mathbf{q}}_i &= \mathbf{Q}^{-1}(\dot{\mathbf{p}}_i - \frac{d\mathbf{Q}}{dt} \mathbf{q}_i) = \mathbf{Q}^{-1}(\dot{\mathbf{p}}_i - [\dot{\mathbf{p}}_1 \quad \dot{\mathbf{p}}_2 \quad \dot{\mathbf{p}}_3] \begin{bmatrix} \mathbf{q}_{1i} \\ \mathbf{q}_{2i} \\ \mathbf{q}_{3i} \end{bmatrix}) \quad (10) \\ &= \mathbf{Q}^{-1}(\dot{\mathbf{p}}_i - \mathbf{q}_{1i} \dot{\mathbf{p}}_1 - \mathbf{q}_{2i} \dot{\mathbf{p}}_2 - \mathbf{q}_{3i} \dot{\mathbf{p}}_3) \end{aligned}$$

According to the IBVS we know the derivative of the current image point is $\dot{\mathbf{p}}_i = \mathbf{J}_i \mathbf{v}_c$. By plugging this into equation (10), we can obtain:

$$\mathbf{J}_{q_i} = \mathbf{Q}^{-1}(\mathbf{J}_i - \mathbf{q}_{1i} \mathbf{J}_1 - \mathbf{q}_{2i} \mathbf{J}_2 - \mathbf{q}_{3i} \mathbf{J}_3), i=1, \dots, n \quad (11)$$

where \mathbf{J}_i is the standard image Jacobian matrix.

$$\mathbf{J}_i = \begin{bmatrix} -\frac{f}{Z_i} & 0 & \frac{x_i}{Z_i} & \frac{x_i y_i}{f} & -f - \frac{x_i}{f} & y_i \\ 0 & -\frac{f}{Z_i} & \frac{y_i}{Z_i} & f + \frac{y_i^2}{f} & \frac{x_i y_i}{f} & -x_i \end{bmatrix} \quad i=1, 2, \dots, n \quad (12)$$

\mathbf{J}_{q_i} can be written just as follows:

$$\mathbf{J}_{q_i} = [\phi_i(Z_1, Z_2, Z_3, Z_i) \quad \psi_i] \begin{bmatrix} \mathbf{K} & 0 \\ 0 & \mathbf{F}(\mathbf{K}) \end{bmatrix} \quad i=1, 2, \dots, n$$

$$\text{where, } \phi_i = \mathbf{Q}^{-1} \begin{bmatrix} \sum_{j=1}^3 \frac{q_{ji}}{Z_i} - \frac{1}{Z_i} & 0 & \frac{u_i}{Z_i} - \sum_{j=1}^3 \frac{u_{ji}}{Z_i} \\ 0 & \sum_{j=1}^3 \frac{q_{ji}}{Z_i} - \frac{1}{Z_i} & \frac{v_i}{Z_i} - \sum_{j=1}^3 \frac{v_{ji}}{Z_i} \\ 0 & 0 & 0 \end{bmatrix}$$

$$\psi_i = \mathbf{Q}^{-1} \begin{bmatrix} u_i v_i - \sum_{j=1}^3 u_j v_j q_{ji} & \sum_{j=1}^3 u_i^2 q_{ji} - u_i^2 & v_i - \sum_{j=1}^3 v_j q_{ji} \\ v_i^2 - \sum_{j=1}^3 v_j^2 q_{ji} & \sum_{j=1}^3 u_i v_i q_{ji} - u_i v_i & \sum_{j=1}^3 u_j q_{ji} - u_i \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{F}(\mathbf{K}) = \begin{bmatrix} \frac{\sin \theta}{fk_v} & \frac{-\cos \theta}{fk_u} & 0 \\ 0 & \frac{1}{fk_u} & 0 \\ 0 & 0 & \frac{1}{k_u k_v} \end{bmatrix}$$

2.3.3 Definition of the error function in the Invariant Space

The error function in the invariant space can be defined as follows:

$$\mathbf{e} = \hat{\mathbf{J}}_q^+ (\mathbf{s} - \mathbf{s}^*) \quad (13)$$

where $\mathbf{s} = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n)$ is n points of the observed object coordinate vector in the invariant space and current position and $\mathbf{s}^* = (\mathbf{q}_1^*, \mathbf{q}_2^*, \dots, \mathbf{q}_n^*)$ is the coordinate vector in the invariant space in the desired position of these points. $\hat{\mathbf{J}}_q^+$ is the pseudo-inverse of the estimated image Jacobian matrix ($3n \times 6$). If the visual servoing law is:

$$\mathbf{u} = -\lambda \mathbf{e} \quad (14)$$

where λ is a positive scalar factor, the robot can be driven back to the desired position only if \mathbf{s} converges to \mathbf{s}^* , meaning that the end operator of the robot arrives in the desired position.

3. Basic principle of the depth adaptive zooming visual servoing

The invariant visual servoing approach described above provides a good methodology to control the robot with a zooming camera. In this section, we will bring the zoom control and depth adaptive estimation into the invariant visual servoing method and proposed a depth adaptive zooming visual servoing method. The purpose of the proposed method is to keep the observed object always staying in the field of view of camera during the visual servoing through change the camera focus length, and to adaptively estimate the depth of the image feature points on the object during the visual servoing.

3.1 The Basic Structure of the Depth Adaptive Zooming Visual Servoing

The whole servoing structure diagram of the proposed robot depth adaptive zooming visual servoing method can be depicted in Fig. 1. The desired input \mathbf{p}^* shows the image feature points on the observed object with respect to the camera in the desired position. The transformed points \mathbf{s}^* are the corresponding features in the invariant image space, which are independent of the internal camera parameters. Similarly, \mathbf{s} are features in the invariant image space of the same image feature points in the current frame. The visual servoing law: $\mathbf{u} = -\lambda \mathbf{e} = -\lambda \hat{\mathbf{J}}_q^+ (\mathbf{s} - \mathbf{s}^*)$ can be calculated with the error in the invariant image space. We could control the robot to the goal position from the invariant space to the robot motion space by this visual servoing law. The robot visual positioning task will be accomplished once the image features in the invariant space are consistent with themselves at the desired position^[16]. The main aim of the

zooming control (see details in subsection 3.2) is to ensure the interest object is lying in the field of view of the camera during the visual servoing. The aim of the depth estimation is to adaptively estimate the depth of the image feature points on the object during the visual servoing (see details in subsection 3.3).

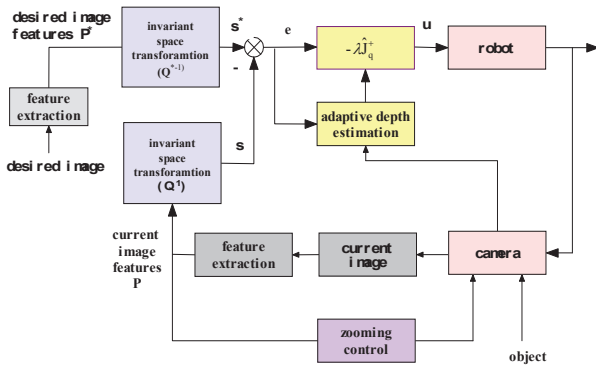


Figure 1. Structure diagram of depth adaptive zooming visual servoing.

3.2 Zooming Control

The purpose of zooming control can be described as follows:

During the robot visual positioning process, zooming out when the object moves away from the camera (even out of the field of view) and zooming in to produce an expanding image of the interest object can achieve high precision positioning [17-19]. Note that the zooming control can't change the resolution of the image captured from the camera, but can change the local resolution of the interest object.

First, we define δ as the distance from the nearest feature point on the object to the border of the image plane. $\delta = \min(u_1, v_1, u_{\max} - u_1, v_{\max} - v_1)$, where u_{\max} and v_{\max} are respectively the image dimensions in the \bar{u} and \bar{v} direction. We can control the zoom in order to keep the nearest point to the image borders at a given distance. The zooming areas in the image plane are shown in Fig. 2.

If the object is in region 1, i.e., δ is in section 1, in order to improve the image resolution of the interest object, we can enlarge the focal length. Zooming means

$$\left. \frac{\partial \delta}{\partial t} \right|_{\delta > \delta_{\max}} > 0.$$

If the object is in the dark grey area, i.e., δ is in region 2, there is no zooming in this area which can keep all the feature points in the view of the camera stably. That is

$$\left. \frac{\partial \delta}{\partial t} \right|_{\delta_{\min} < \delta < \delta_{\max}} = 0.$$

If the object is in the light grey area, i.e., δ is in region 3, it means that the object will escape from the field of view of the camera, so we should decrease zoom in this area to obtain a sufficient field of view of the camera. The corresponding mathematical expression can be written as

$$\left. \frac{\partial \delta}{\partial t} \right|_{\delta < \delta_{\min}} < 0.$$

According to the different regions where the object is, we can adopt the corresponding zooming control rules to improve the image resolution on the premise of guaranteeing all the features on the object stay in the field of view of the camera.

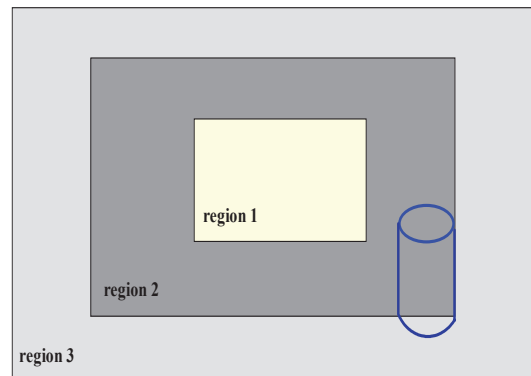


Figure 2. Zooming areas.

3.3 Depth Adaptive Estimation

As the description in subsection 2.3 states, the visual servoing law is $\mathbf{u} = -\lambda \mathbf{e} = -\lambda \hat{\mathbf{J}}_q^+ (\mathbf{s} - \mathbf{s}^*)$ and \mathbf{J}_q is an image Jacobian matrix with relation to the target depth, which is a time-varying parameter. Depth estimation is a challenging problem in computer vision and robotics. Many studies have presented methods to estimate the unknown depth distance in robot motion in the 3D environment [14, 15]. However, these depth estimation methods are only suitable for fixed camera intrinsic parameters. In this paper, we extend these depth estimation methods to zooming visual servoing, where the camera intrinsic parameter changes during the visual servoing. We propose a depth adaptive estimation method in the invariant space. The basic steps of this method can be described as follows:

Step 1: to transform the image features in the original image space into the ones in the invariant space by using (6) and (7).

Step 2: to choose a state function not only related to the feature error, but also to the depth parameter.

Step 3: to design a nonlinear depth adaptation scheme according to Lyapunov stability and prove its rationality. Next we will depict this depth adaptive estimation method in the invariant space in detail.

We define the function of the whole state $(\tilde{\mathbf{s}}, \tilde{\boldsymbol{\varepsilon}})$ as follows:

$$V(\tilde{\mathbf{s}}, \tilde{\boldsymbol{\varepsilon}}) = \frac{1}{2} \tilde{\mathbf{s}}^T \tilde{\mathbf{s}} + \frac{1}{2\gamma_1} \tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}} \quad (15)$$

where $\tilde{\mathbf{s}} = \mathbf{s} - \mathbf{s}^*$ and the definitions of \mathbf{s} and \mathbf{s}^* are described in subsection 2.3. The definition of the error vector $\tilde{\boldsymbol{\varepsilon}}$ is $\tilde{\boldsymbol{\varepsilon}} = [\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n]^T$, $i=1, \dots, n$. Where $\tilde{\varepsilon}_i = \varepsilon_i - \hat{\varepsilon}_i$, ε_i is the reciprocal of the target depth $\varepsilon_i = 1/Z_i$ and $\hat{\varepsilon}_i$ is the estimated value of ε_i .

In order to ensure system stability, we must address the system satisfy Lyapounov theories.

Observing the matrix \mathbf{J}_q , the depth has a proportional relationship with the first three columns of the matrix.

By expanding equation (15), we can observe:

$$V(\tilde{\mathbf{s}}, \tilde{\boldsymbol{\varepsilon}}) = \frac{1}{2} \sum_{i=1}^n \tilde{q}_{1i}^2 + \tilde{q}_{2i}^2 + \tilde{q}_{3i}^2 + \frac{1}{\gamma_1} \tilde{\varepsilon}_i^2 \quad (16)$$

The derivative of equation (16) is:

$$\begin{aligned} \dot{V} &= \sum_{i=1}^n (\tilde{q}_{1i} \dot{\tilde{q}}_{1i} + \tilde{q}_{2i} \dot{\tilde{q}}_{2i} + \tilde{q}_{3i} \dot{\tilde{q}}_{3i} + \frac{1}{\gamma_1} \tilde{\varepsilon}_i \dot{\tilde{\varepsilon}}_i) \\ &= \sum_{i=1}^n ((\tilde{q}_{1i} (\mathbf{Q}^{-1} \boldsymbol{\phi}_i \mathbf{K})|_{r1} \mathbf{V}_T + (\mathbf{Q}^{-1} \boldsymbol{\psi}_i \mathbf{F}(\mathbf{K}))|_{r1} \mathbf{V}_R \\ &\quad + \tilde{q}_{2i} ((\mathbf{Q}^{-1} \boldsymbol{\phi}_i \mathbf{K})|_{r2} \mathbf{V}_T + (\mathbf{Q}^{-1} \boldsymbol{\psi}_i \mathbf{F}(\mathbf{K}))|_{r2} \mathbf{V}_R) \\ &\quad + \tilde{q}_{3i} ((\mathbf{Q}^{-1} \boldsymbol{\phi}_i \mathbf{K})|_{r3} \mathbf{V}_T + (\mathbf{Q}^{-1} \boldsymbol{\psi}_i \mathbf{F}(\mathbf{K}))|_{r3} \mathbf{V}_R) \\ &\quad + \frac{1}{\gamma_1} (\tilde{\varepsilon}_i \dot{\hat{\varepsilon}}_i - \varepsilon_i \dot{\hat{\varepsilon}}_i - \tilde{\varepsilon}_i \dot{\hat{\varepsilon}}_i)) \\ &= \sum_{i=1}^n ((\varepsilon_i (\tilde{q}_{1i} ((\mathbf{Q}^{-1} \mathbf{C}_i \mathbf{K}))|_{r1} + \tilde{q}_{2i} ((\mathbf{Q}^{-1} \mathbf{C}_i \mathbf{K}))|_{r2} \\ &\quad + \tilde{q}_{3i} ((\mathbf{Q}^{-1} \mathbf{C}_i \mathbf{K}))|_{r3}) \mathbf{V}_T - \frac{\hat{\varepsilon}_i}{\gamma_1} \\ &\quad + (\mathbf{Q}^{-1} \boldsymbol{\psi}_i \mathbf{F}(\mathbf{K}))|_{r1} \mathbf{V}_R + (\mathbf{Q}^{-1} \boldsymbol{\psi}_i \mathbf{F}(\mathbf{K}))|_{r2} \mathbf{V}_R \\ &\quad + (\mathbf{Q}^{-1} \boldsymbol{\psi}_i \mathbf{F}(\mathbf{K}))|_{r3} \mathbf{V}_R + \frac{1}{\gamma_1} (\hat{\varepsilon}_i \dot{\tilde{\varepsilon}}_i - \tilde{\varepsilon}_i \dot{\hat{\varepsilon}}_i)) \end{aligned}$$

If we use the depth adjusting law:

$$\dot{\hat{\varepsilon}}_i = \gamma_1 (\mathbf{Q}^{-1} (\mathbf{C}_i \mathbf{K}) \mathbf{v}_T)^T \tilde{\mathbf{q}}_i - \gamma_2 (\mathbf{Q}^{-1} (\mathbf{C}_i \mathbf{K}) \mathbf{v}_T)^T \mathbf{e}_i \quad (17)$$

Where, $\gamma_1 > 0, \gamma_2 > 0, i=1, 2, \dots, n$, $\mathbf{e}_i = \hat{\mathbf{d}}_i - \mathbf{d}_i^*$ is called as prediction error, $\mathbf{d}_i = \mathbf{Q}^{-1} (\mathbf{C}_i \mathbf{K}) \mathbf{v}_T \varepsilon_i$, $i=1, 2, \dots, n$. $\dot{\boldsymbol{\varepsilon}} = [\dot{\varepsilon}_1, \dots, \dot{\varepsilon}_n]^T$ is the derivative of the reciprocal of the depth parameters and $\tilde{\boldsymbol{\varepsilon}} = [\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n]^T$, $i=1, \dots, n$ is the error of the reciprocal of the depth parameters.

So equation (17) can be simplified as follows:

$$\dot{V} = \sum_{i=1}^n \tilde{\mathbf{q}}_i^T \hat{\mathbf{J}}_q \mathbf{V}_C - \frac{1}{\gamma_1} \tilde{\varepsilon}_i \dot{\hat{\varepsilon}}_i - \frac{\gamma_2}{\gamma_1} \left\| \mathbf{Q}^{-1} \mathbf{C}_i \mathbf{K} \mathbf{V}_T \right\|^2 \tilde{\varepsilon}_i^2 \quad (18)$$

where:

$$\mathbf{C}_i = \begin{bmatrix} \sum_{j=1}^3 q_{ji} - 1 & 0 & \mathbf{u}_i - \sum_{j=1}^3 \mathbf{u}_{ji} \\ 0 & \sum_{j=1}^3 q_{ji} - 1 & \mathbf{v}_i - \sum_{j=1}^3 \mathbf{v}_{ji} \\ 0 & 0 & 0 \end{bmatrix}$$

The camera velocity \mathbf{v}_c , relative to the target, can be describe as $\mathbf{v}_c = \begin{bmatrix} \mathbf{v}_T^T & \mathbf{v}_R^T \end{bmatrix}^T$, where $\mathbf{v}_T = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]^T$ is the translation velocity vector of the camera and $\mathbf{v}_R = [\mathbf{v}_4 \ \mathbf{v}_5 \ \mathbf{v}_6]^T$ is the rotation velocity vector of the camera.

If we define the vector $\hat{\mathbf{y}} = \hat{\mathbf{J}}_q^T \tilde{\mathbf{s}} \in \mathfrak{R}^6$, equation (18) can be reduced to:

$$\begin{aligned} \dot{V} &= -\lambda \tilde{\mathbf{s}}^T \hat{\mathbf{J}}_q \hat{\mathbf{J}}_q^T \tilde{\mathbf{s}} - \frac{1}{\gamma_1} \tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}} - \frac{\gamma_2}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}}_y \right\|^2 = -\lambda \hat{\mathbf{y}}^T (\hat{\mathbf{J}}_q^T \hat{\mathbf{J}}_q)^{-1} \hat{\mathbf{y}} - \frac{1}{\gamma_1} \tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}} - \frac{\gamma_2}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}}_y \right\|^2 \\ &\leq -\frac{\lambda}{\hat{\boldsymbol{\sigma}}_M^2} \left\| \hat{\mathbf{y}} \right\|^2 + \frac{1}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}} \right\| \left\| \dot{\boldsymbol{\varepsilon}} \right\| - \frac{\gamma_2}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}}_y \right\|^2 \end{aligned}$$

where $\hat{\boldsymbol{\sigma}}_M = \max(\hat{\mathbf{J}}_q)$, $\hat{\boldsymbol{\sigma}}_m = \min(\hat{\mathbf{J}}_q)$ and

$$\tilde{\boldsymbol{\varepsilon}}_y = \left[\left\| \mathbf{Q}^{-1} \mathbf{C}_i \mathbf{K} \mathbf{V}_T \right\| \tilde{\varepsilon}_1, \dots, \left\| \mathbf{Q}^{-1} \mathbf{C}_i \mathbf{K} \mathbf{V}_T \right\| \tilde{\varepsilon}_n \right]^T$$

We know that $V \geq 0$. If $\dot{V} \leq 0$ in a region then the system with the adaptive estimation method is locally asymptotically stable. If:

$$\frac{1}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}} \right\| \left\| \dot{\boldsymbol{\varepsilon}} \right\| \leq \frac{\lambda}{\hat{\boldsymbol{\sigma}}_M^2} \left\| \hat{\mathbf{y}} \right\|^2 + \frac{\gamma_2}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}}_y \right\|^2 \quad (19)$$

where $\left\| \tilde{\boldsymbol{\varepsilon}}_y \right\| \geq \delta \left\| \tilde{\boldsymbol{\varepsilon}} \right\|$, $\delta = \min(\mathbf{C}_i \mathbf{V}_T) > 0$. The \dot{V} is negative in the region:

$$\mathfrak{R} = \{ \tilde{\mathbf{s}} \in \Omega, \tilde{\boldsymbol{\varepsilon}} \in \mathfrak{R}^n : \frac{1}{\gamma_1} \left\| \tilde{\boldsymbol{\varepsilon}} \right\| \left\| \dot{\boldsymbol{\varepsilon}} \right\| \leq \frac{\lambda}{\hat{\boldsymbol{\sigma}}_M^2} \left\| \tilde{\boldsymbol{\varepsilon}} \right\|^2 \}$$

meaning that the state error $\tilde{\mathbf{s}}$ converges to the residual

$$\text{set } B(\tilde{\mathbf{s}}, \frac{\xi \hat{\boldsymbol{\sigma}}_M}{2\delta \hat{\boldsymbol{\sigma}}_m \sqrt{\lambda \gamma_1 \gamma_2}})$$

As the unknown object depth, if the visual servoing law is $\mathbf{u} = -\lambda \mathbf{e} = -\lambda \hat{\mathbf{J}}_q^+ (\mathbf{s} - \mathbf{s}^*)$, the adaptive adjustment mechanism can guarantee that the states of its system are uniformly bounded.

4. Simulation Results

In this section we build a robot depth adaptive zooming visual servoing simulation model in a MATLAB environment, based on Peter's robotics toolbox [20] and conduct three kinds of robot visual positioning simulation experiments to validate the performance of the proposed depth adaptive zooming visual servoing.

The first simulation experiment aims to demonstrate the rationality of the whole structure with zooming control. The second simulation involves plugging into the depth adaptive method and is in contrast with the first simulation, in order to validate the visual servoing performance of the proposed approach. The third simulation aims to demonstrate the validity of the proposed approach in a different initial position.

4.1 Related Parameter Settings

The camera is in the eye-in-hand configuration. The related parameters used in the simulation experiments are as follows:

- the number of the feature points $n=12$
- coordinates of the principle point $u_0=256, v_0=256$ (in pixels)
- magnifications in the direction \bar{u} and \bar{v} $k_u=20000, k_v=20000$ (in pixels/m)
- angle between \bar{u} and \bar{v} axes" $\theta=90^\circ$
- initial focal length $f_0=0.015$ (in meters)
- desired focal length $f_d=0.012$ (in meters)
- coordinates of the 12 image points in the 3D space with different depth points=

[0.5 -0.5 2.01; 0.9 0.2 2.04; -0.2 -0.2 2;
0.9 -0.5 2.01; 0.9 -0.9 2.01; 0.5 -0.9 2.01;
0.7 -0.7 2.01; 0.5 -0.2 2.04; 0.7 0 2.04;
0.2 0.2 2; 0.2 -0.2 2; -0.2 0.2 2],

- the initial camera pose

Tcamera_0= [0 0.0001 1 0.5;
 -0.0001 1 -0.0001 -0.3;
 -1 -0.0001 0 0.6;
 0 0 0 1]

Suppose the robot move 700mm in the X-minus direction, 300mm in the Y direction, 100mm in the Z-minus direction and take the rotation -0.3rad along the Z direction. Then we get the goal position:

Tcamera_e= [0 0.0001 1 0.4;
 -0.2956 0.9553 -0.0001 0;
 -0.9553 -0.2956 0 -0.1;
 0 0 0 1],

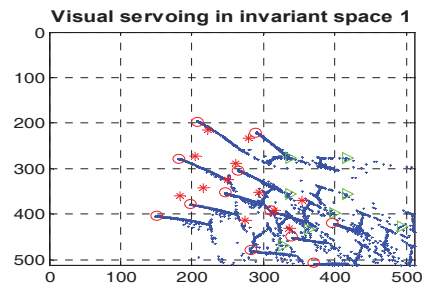
The visual servoing law used is $\mathbf{u} = -\lambda \mathbf{e} = -\lambda \hat{\mathbf{J}}_q^+ (\mathbf{s} - \mathbf{s}^*)$ where the control gain $\lambda=0.3$ and the image Jacobian matrix $\hat{\mathbf{J}}_q$ can be computed by (11).

We don't consider the depth estimation in the first simulation experiment, but consider the depth estimation, with the depth adjusting gains $\gamma_1=10^2, \gamma_2=10^3$ in the second simulation.

4.2 Results and Discussion

4.2.1 Simulation experiment 1: robot zooming visual servoing without depth estimation

In this simulation experiment, 12 image feature points on the target object at the desired position and at the initial position are obtained respectively. Then the robot moves from the initial position to the desired position using zooming visual servoing without depth estimation. The motion trajectory of the 12 feature points in the image plane is shown in Fig. 3. At the very beginning of the servoing, the focal length is 12mm and there are four points outside of the field of view. The camera zooms out in order to assure the interest object points are all in view at the beginning of the servoing. During the late servoing, zooming in improves the image resolution and the final focus value is 17.3mm (Fig. 5 (b)). The simulation results show that the control method can make the error of characteristic vectors in invariant space of the target object zero (Fig. 5 (a)). As a consequence, the positioning error in directions X, Y and Z are Xerror=1.66mm, Yerror=0.59mm, Zerror=6.44mm, respectively. The rotation error converges to 0 (Fig. 4 (a) and Fig. 4 (b)). So the robot can reach the desired position.



(the green triangle points represent the initial position, the red asterisk ones represent the desired position, the red circle ones represent the real position after the servoing)

Figure 3. The image points trajectory (blue points) without depth estimation.

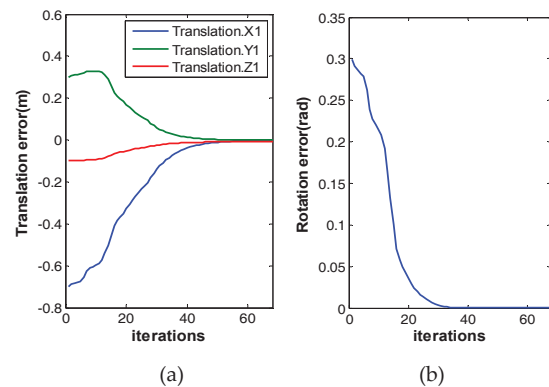


Figure 4. Visual positioning error without depth estimation; (a) Translation error. (b) Rotation error around Z axes.

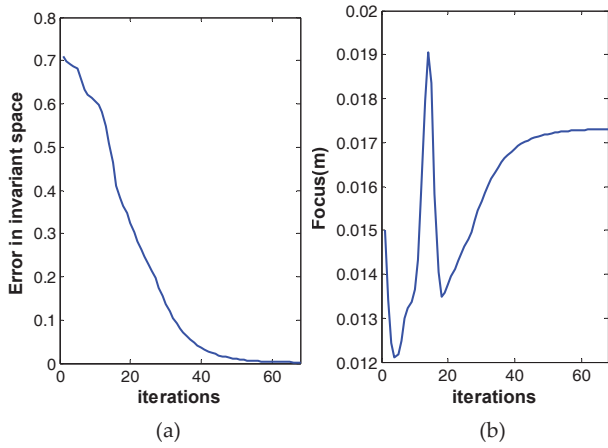
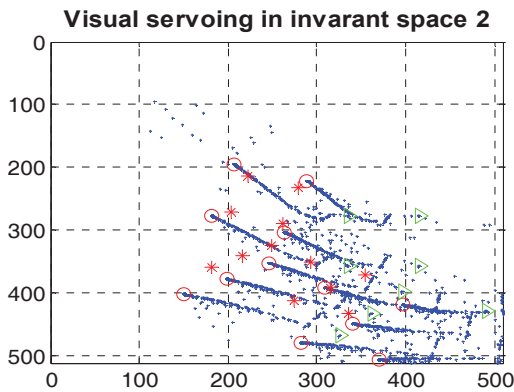


Figure 5. Error in the invariant space and change of focal length without depth estimation; (a) The total error in the invariant space. (b) Focal length changes in visual servoing.

4.2.2 Simulation experiment 2: robot zooming visual servoing with depth adaptive estimation

In this section, we conducted another robot 4DOF visual positioning simulation experiment using zooming visual servoing with a depth adaptive estimation method. The experiment parameters used in this experiment are the same as the ones used in the first simulation experiment. The visual positioning results using zooming visual servoing with a depth adaptive estimation are shown in Fig. 6 - Fig. 8. The motion trajectory of the 12 feature points in the image plane is shown in Fig. 6.



(the green triangle points are at the initial position, the red asterisk ones are at the desired image from the learning, the red circle ones are the image points at the real position after the servoing)

Figure 6. The image points trajectory (blue points) with depth estimation

The errors of the characteristic vectors in the invariant space of the target object are zero (Fig. 8 (a)) and the focal lengths finally converge to 17.1mm (Fig.8 (b)). The errors along the X, Y and Z axes are $X_{error}=0.29\text{mm}$, $Y_{error}=0.35\text{mm}$ and $Z_{error}=2.59\text{mm}$. The rotation errors converge to 0 and are shown in Fig.7 (a) and Fig.7 (b) respectively. The simulation results show that the positioning error along the Z axis is smaller when depth

estimation is considered in the zooming visual servoing. The whole depth converge process is shown in Fig. 10. Comparison results of the convergence performance of the two zooming visual servoing methods are shown in Fig. 9. It can be seen from the Fig. 9 that the zooming visual servoing with a depth adaptive estimation method has a better convergence performance.

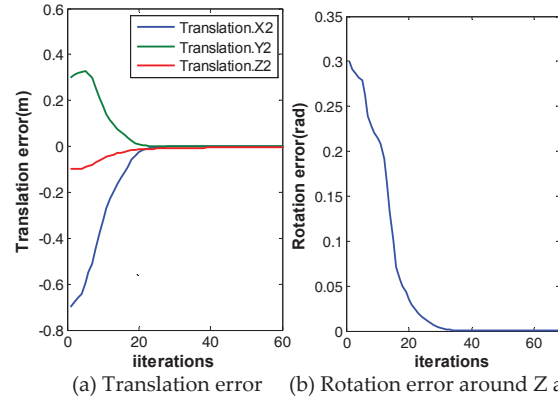


Figure 7. Visual positioning error with depth estimation

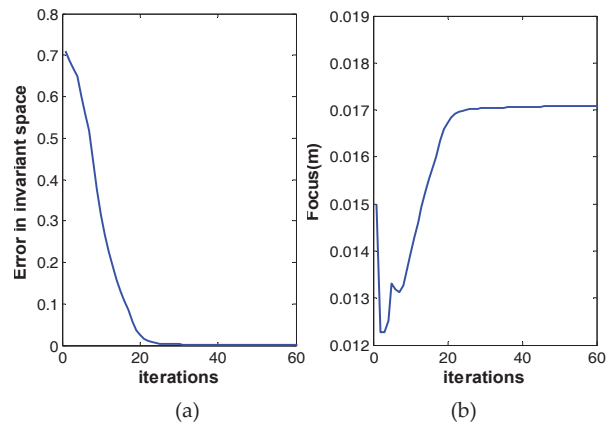


Figure 8. Error in the invariant space and change of focal length with depth estimation; (a) The total error in the invariant space; (b) Focal length changes in visual servoing.

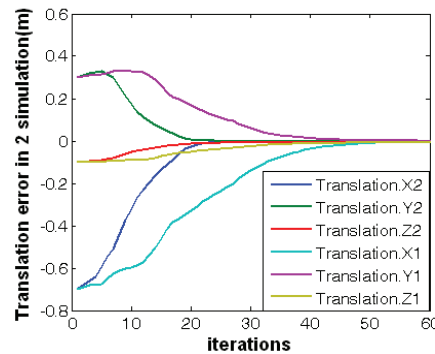


Figure 9. Comparison of the convergence of the two zooming visual servoing method

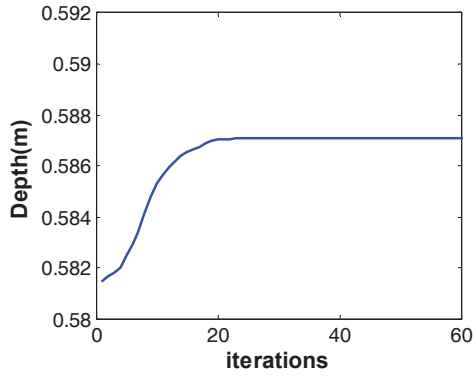


Figure 10. Depth convergence curve

	Error(mm)			Run time(s) (in MATLAB)
	X	Y	Z	
Simulation 1	1.66	0.59	6.44	22.84
Simulation 2	0.29	0.35	2.59	21.82

Table 1. The translation convergence errors along X, Y, Z axes in the two simulations experiments.

4.2.3 Simulation experiment 3: robot zooming visual servoing with depth adaptive estimation in the different initial position.

In order to demonstrate the validity of the proposed approach further. We conduct another robot 4DOF visual positioning simulation experiment with a different initial camera pose.

Suppose the robots moves 300mm in the X-minus direction, 200mm in the Y direction and -200mm in the Z-minus direction.

Then we get the new initial camera pose:

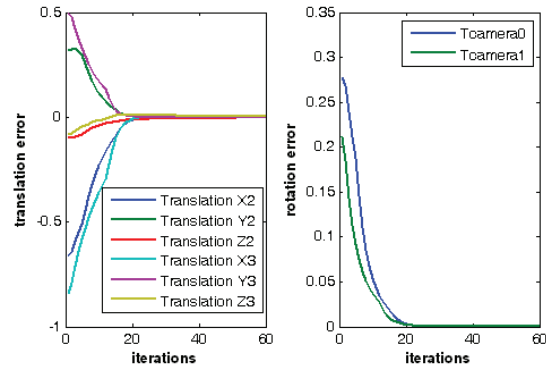
$$T_{camera_1} = \begin{bmatrix} 0 & 0.0001 & 1 & 0.7; \\ -0.0001 & 1 & -0.0001 & -0.5; \\ -1 & -0.0001 & 0 & 0.3; \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Other experimental parameters used in this experiment are the same as the ones used in the first two simulation experiments.

The visual positioning results using zooming visual servoing with depth adaptive estimation in the different initial camera position are shown in Fig.11 - Fig.13.

The translation errors along the X, Y and Z axes and the rotation error around the Z axis are shown in Fig. 11 (a) and Fig. 11 (b) respectively. The errors of characteristic vectors in the invariant space of the target object are zero (Fig. 12 (a)) and the focal length finally converge to 17.1mm in Tcamera_0 and 23.0mm in Tcamera_1 (Fig. 12 (b)). The whole depth convergence process is shown in Fig. 13. The simulation results show that our approach

has good visual positioning performance in the different camera initial positions.



(a) Translation error (b) Rotation error around Z axis.

Figure 11. Visual positioning error in the different initial camera position

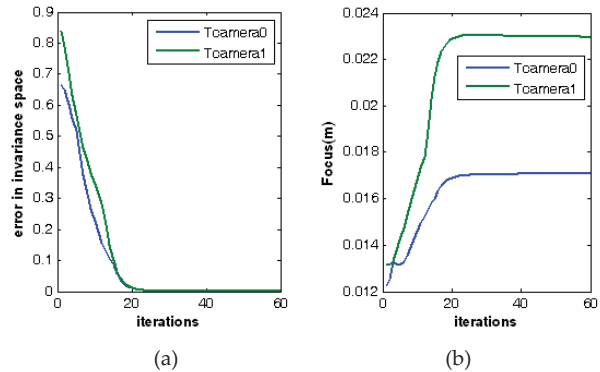


Figure 12. Error in the invariant space and change of focal length in the different initial camera position; (a) The total error in the invariant space; (b)Focal length changes in visual servoing.

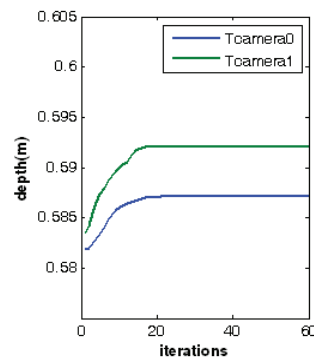


Figure 13. Depth convergence curve in the different initial camera position

Error(mm)	X	Y	Z
Simulation 2	0.29	0.35	2.59
Simulation 3	0.49	0.31	3.61

Table 2. The translation convergence errors along X, Y, Z axes in the two different camera initial positions.

The simulation results show that the zoom control method can be used in the invariant visual servoing

system. The purpose of the zooming control is to extend the camera's field of view when the target is moving out of range and improve the local resolution of the object of interest when the target is in range of the camera view. In addition, the proposed depth adaptive zooming visual servoing method can guarantee closed-loop stabilization to achieve the 4-DOF robot visual positioning task. It can be seen from Table 1, in contrast to simulation 1, the second simulation system with a depth adaptive estimation has higher visual servoing precision and the whole state of the robot visual servoing system is bounded, though the depth converges to a stable value instead of the true value.

5. Conclusion

In this paper, a depth adaptive zooming visual servoing method is proposed to solve view visibility problems. In order to enlarge the field of view, the visual servoing system in the invariant space with zooming control is performed. This could improve the accuracy of image resolution of the target. In order to overcome the defects of the constant depth method, a nonlinear depth adaptive estimation method for robot zooming visual servoing is proposed. With robustness in the invariant space, this adaptive method can improve the accuracy of the servoing system under the zooming condition. The adaptive adjustment mechanism can guarantee that the states of its system are uniformly bounded. The simulation results of robot 4DOF visual positioning show that the proposed method has a better convergence performance for the image error.

No specific feature extraction algorithm is considered in the simulation experiment and some simulation points are defined as the image feature points. The performance of the feature point's extraction will affect the final robot visual servoing. In some cases, the whole robot visual servoing would become unstable, if there is a big mistake while extracting these points. Changing lighting conditions would affect the performance of the feature point's extraction. In addition to a robot/camera system with eye-in-hand configuration, the image captured by the camera will undergo some deformations because of the change in robot pose and camera focus length during the robot zooming visual servoing. These deformations can be locally well approximated by affine transformations of the image plane. In the future, we will apply the affine invariant feature extraction algorithm^[21, 22] to extract the affine invariant feature points on the object and extend the proposed robot depth adaptive zooming visual servoing method to real nature scenes .

6. Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No. No. 61203345, No. 61174101, the Shaanxi Provincial Natural Science

Foundation of China under Grant No. 2009JQ8011 and Educational Commission of Shaanxi Province of China under Grant No. 2010JK737.

7. References

- [1] Chaumette F & Hutchinson S. (2006) Visual Servo Control, Part I: Basic Approaches. *IEEE Robotics & Automation Magazine* 13(4):82-90.
- [2] Fayman, J. A., Sudarsky, O & Rivlin, E. (1998) Zoom tracking. *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*. Leuven, Belgium: ,2783-2788
- [3] Fang, Y. C. (2008) A survey of robot visual servoing. *CAAI transaction on intelligent systems* 3(2):109-113
- [4] Mezouar, Y. & Chaumette, F. (2000) Path planning in image space for robust visual servoing. *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, San Francisco: 2759–2764.
- [5] Schramm F. & Morel G. (2006).Ensuring visibility in calibration-free path planning for image-based visual servoing. *IEEE Transaction on Robotics* 22(4): 848-854
- [6] Yao Z. & Gupta K. (2007) Path planning with general end-effector constraints. *Robotics and Autonomous Systems* 55(4): 316-327
- [7] Chesi, G. (2009) Visual Servoing Path Planning via Homogeneous Forms and LMI Optimizations. *IEEE Transactions on Robotics* ,25(2): 281-291
- [8] Deng L., Janabi-Sharifi F. & William J. (2005) Hybrid Motion Control and Planning Strategies for Visual Servoing. *IEEE Transactions on Industrial Electronics* 52(4): 1024-1040.
- [9] Kazemi, M., Gupta, K. & Mehrandezh M. (2010) Path Planning for Visual Servoing: A Review and Issues. *Visual Servoing via Advanced Numerical Methods*, *Lecture Notes in Control and Information Sciences* 401: 189-207
- [10] Tordoff, B. J. & Murray, D. W. (2007) A method of reactive zoom control from uncertainty in tracking. *Computer Vision and Image Understanding*: 131-144
- [11] Kumar, S., Micheloni, C. & Piciarelli, C. (2009) Stereo Localization Using Dual PTZ Cameras. *Proceeding of IEEE International Conference on Computer Analysis of Images and Patterns*, Munster, Germany
- [12] Malis E. (2004) Visual servoing invariant to changes in camera-intrinsic Parameters. *IEEE Transaction on Robotics and Automation* 20(1): 72-81.
- [13] Malis, E. & Benhimane, S. (2003) Vision-based control with respect to planar and non-planar objects using a zooming camera. *Proceeding of IEEE International Conference on Advanced Robotics*, Coimbra, Portugal.
- [14] Conticelli, F. & Allotta, B. (2001) Nonlinear Controllability and Stability Analysis of Adaptive Image-Based Systems. *IEEE Transactions on Robotics and Automation* 17(2):208-214.

- [15] Zachi ,A. R. L., Liu, H. & Lizarralde, F. (2004) Performing stable 2D adaptive visual positioning/tracking control without explicit depth measurement. Proceedings of the 2004 IEEE International Conference on Robotics and Automation 3: 2297 – 2302
- [16] Su, J. B. (2010) Uncalibrated robotic hand-eye coordination. Beijing: Electronic industry press
- [17] Hatano, K. & Hashimoto, K (2003) Image-Based Visual Servo Using Zoom Mechanism. SICE Annual Conference, Fukui,
- [18] Benhimane, S. & Malis, E. (2003) Vision-based control with respect to planar and non-planar objects using a zooming camera. The 11th International Conference on Advanced Robotics, Coimbra, Portugal
- [19] Radchenko, O., Pape, A. & Graser, A. (2003) Visual servoing with adjustable zoom cameras. The 8th International Conference on Rehabilitation Robotics: 22-25
- [20] Corke, P. J. (2007) MATLAB toolboxes: robotics and vision for students and teachers, IEEE Robotics and Automation Magazine 14(4): 16-17
- [21] Morel, J. & Yu, G. (2009) ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences 2(2): 438-469
- [22] Yu, G. & Morel J. (2011) ASIFT: An Algorithm for Fully Affine Invariant Comparison. Image Processing On Line: 47-52