Journal of Intelligent & Robotic Systems (2022) 104: 10
https://doi.org/10.1007/s10846-021-01516-w

REGULAR PAPER

Visual Servoing of Unknown Objects for Family Service Robots

Jing Xin [1] · Caixia Dong [1] · Youmin Zhang [2] · Yumeng Yao [1] · Ailing Gong [1]

Received: 25 July 2020 / Accepted: 11 October 2021 / Published online: 21 December 2021

Abstract
Aiming at satisfying the increasing demand of family service robots for housework, this paper proposes a robot visual servoing scheme based on the randomized trees to complete the visual servoing task of unknown objects in natural scenes. Here, "unknown" means that there is no prior information on object models, such as template or database of the object. Firstly, an object to be manipulated is randomly selected by user prior to the visual servoing task execution. Then, the raw image information about the object can be obtained and used to train a randomized tree classifier online. Secondly, the current image features can be computed using the well-trained classifier. Finally, the visual controller can be designed according to the error of image feature, which is defined as the difference between the desired image features and current image features. Five visual positioning of unknown objects experiments, including 2D rigid object and 3D non-rigid object, are conducted on a MOTOMAN-SV3X six degree-of-freedom (DOF) manipulator robot. Experimental results show that the proposed scheme can effectively position an unknown object in complex natural scenes, such as occlusion and illumination changes. Furthermore, the developed robot visual servoing scheme has an excellent positioning accuracy within 0.05 mm positioning error.

Keywords Robot visual servoing · Natural scenes · Randomized tree classifier · Unknown objects

1 Introduction

Accurate positioning of objects is the crux to real-world applications of family service robots. Obviously, the basic function of service robots is that it can operate, grasp and move a specific object selected by the user freely. The first step to implement the above task is to make a robot positioning an object with high accuracy and strong robustness. At present, most visual servoing approaches, which have been widely and successfully used for applications where the object to be manipulated is known beforehand [1–4] or the scene is known beforehand [5]. Unlike industrial robots, family service robots usually work in highly unstructured environments [6, 7], and need more intelligence than industrial robots to perform given tasks [8]. In this environment, the object may be "unknown", and accurate positioning becomes a more challenging problem. This "unknown" means that no assumption is made on the scene structure surrounding the object and no template or database of the object is known to the robot prior to the task execution.

With the new development of science and technology, many researchers have proposed many robot visual servoing methods that can be applied to unknown objects in natural scenes [9–11]. For example, an adaptive visual servoing method based on contour features is proposed in [12], which can be used to recognize and obtain the current position information of the object by learning the contour feature of the object and complete the robot visual servoing task for unknown object. The visual serving method does not require any prior knowledge about the position of the object. However, it does not consider the non-rigid shape change of the object in the visual servoing process. As a result, these methods are only for the visual servoing of rigid bodies, and not applied to the visual servoing of non-rigid bodies.

Visual servoing for non-rigid bodies has always been a difficult and challenging problem because it is difficult to estimate the deformation properties of non-rigid bodies. The research on non-rigid body visual servoing mainly focuses on the representation of the target model. Traditionally, the problem of representing the target model is tackled by establishing its model and estimating its parameters [13–17]. However,

✉ Youmin Zhang
youmin.zhang@concordia.ca

Jing Xin
xinj@xaut.edu.cn

[1] Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing, Xi'an University of Technology, Xi'an 710048, China

[2] Department of Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, QC H3G 1M8, Canada

these methods need to acquire prior knowledge of the properties of these non-rigid bodies in advance. As a result, they are unable to handle unknown objects.

In view of the complexity of non-rigid body model establishment, some researchers have recently proposed methods based on model-free approaches [18–23]. The main idea of these methods is representing the target through artificial marking, which simplifies the extraction of image feature. However, these methods are labor intensive and not efficient. With the development of deep learning approaches, some researchers have proposed to use deep learning algorithms [19, 20] to identify the category information of unknown objects and obtain the best grasping position of the object. However, these methods need to obtain the prior knowledge of the category information of each unknown object by off-line learning, and do not consider the problem of robot visual servoing control of unknown objects in complex natural scenes (such as occlusion and illumination changes). As a result, these methods fail to apply for natural scenes.

In this work, we propose a robot visual servoing scheme based on the randomized trees with an effort for avoiding above-mentioned shortcomings. The main objective and feature of the proposed approach are to complete the visual servoing task of unknown objects in various complex natural scenes towards more practical solutions and applications. As opposed to prior work which uses prior knowledge of objects to establish their model [13–17], the proposed method in this work does not need any prior information on object models, such as template or database of the object. User can randomly select an object to be manipulated prior to the visual servoing task execution. Furthermore, inspired and improved from some deep learning algorithms [19, 20], the proposed method can perform well in complex natural scenes (such as occlusion and illumination changes), which is common in manipulation domains.

The main contributions and novelty of this work can be outlined as follows:

- A visual servoing method based on the randomized trees for unknown objects in natural scenes is proposed. It does not need to acquire any knowledge of the object template and its natural scene in advance, and there are always many objects on the robot operating platform during the positioning process. “Unknown Objects” are randomly specified according to the needs of users, then before the robot servoing task starts, the required data only can be acquired on-line for the specified object, that is, the prior knowledge of the geometric model of the object is not needed in advance.
- The proposed scheme has been tested and evaluated on a real MOTOMAN-SV3X six degree-of-freedom (6-DOF) manipulator robot. Experimental results show that the developed scheme can effectively position an unknown rigid object or non-rigid object in many challenging nature scenes with occlusion and illumination changes and with a high positioning accuracy.

The remainder of this paper is organized as follows. After discussing related work in Section 2, a brief description on the system structure is given in Section 3. The specific implementation of the three main parts of the system, including the construction of the randomized tree classifier, the computation of the current image features, and the design of the visual controller, is described in detail in Section 4. Detailed experimental results and analyses are provided in Section 5. Finally, summary and future works are presented in Section 6.

2 Related Work

Research on the visual servoing of objects (including rigid bodies and non-rigid bodies) can be roughly classified into model-based methods and model-free methods. In model-based methods, the core of these methods is to build the model of the object and estimate its parameters [11, 13–17, 24, 25]. Gratal et al. [11] propose a virtual visual servoing based on saliency map to achieve the visual servoing of unknown objects in natural scenes. Using the approaches in [25, 26], a robot can pick up or place a random object on the desktop, but it cannot operate on a specified object and can only remove all objects one by one on the desktop, which would possibly make the ornament to be cleaned up mistakenly. Jadav et al. [15] used a comprehensive dynamic equation to express the movement of the system to manipulate the deformed object, and then used multiple manipulators (or a claw with multiple fingers) to change the shape of the deformed object. The movement of each manipulator proposes an optimization scheme. This paper verifies the effectiveness of the proposed method through two sets of simulation experiments. However, this method heavily relies on model establishment and parameter estimation for non-rigid body, so it is not suitable for unknown objects.

In model-free methods, the key is the representation of targets. Classical methods represent objects through artificial marking [18–23]. Using the approach in [18], a robot can complete the task of picking up and placing the unknown object by marking the grasping position of the specified object in advance. But the approach is very difficult to apply to realistic unknown natural environments in the family services. Newer methods address the challenge by using deep learning algorithms. A deep learning algorithm [19, 20] is used to recognize the category information of unknown objects and obtain the optimal grasping position of the objects. The approach does not need to obtain the geometric model information of the object prior to the visual servoing task execution.

However, the approach needs to acquire the prior knowledge of the category information of each unknown object by off-line learning, and does not consider the problem of the robot visual servoing control for the unknown object in complex natural scenes (e.g. occlusion and illumination changes).

In this work, to solve the problems of model-based methods and model-free methods, we propose a robot visual servoing scheme under unknown objects in various complex natural scenes. Experimental results show that the proposed scheme can effectively position an unknown object in complex natural scenes with strong robustness to occlusion and illumination variations and small positioning error within 0.05 mm. It is showed that the proposed novel visual servoing scheme can further improve flexible operations of the visual servoing.

3 Problem Description and Overview of System Structure

Robot visual servoing is intended to control the relative pose of robot and object using the visual information. It can allow the robot to work in dynamic and uncontrolled environments [26]. There are two kinds of robot visual servoing structures, including position-based visual servoing (PBVS) and image-based visual servoing (IBVS). This paper adopts the imaged-based visual servoing structure and eye-in-hand configuration. In this configuration, the camera is mounted on the robot end-effector so that it could be moved along with the robot. The relative pose of the robot and object is represented as the difference between the current image features and the desired image feature, then the task of robot visual servoing can be defined as minimizing the relative pose of the robot and object. In other words, under this eye-in-hand configuration and visual servoing structure, both robot visual positioning and visual tracking can be viewed as a positioning problem in an image feature space. Therefore, how to detect an object in complex scenes, to calculate the feature of object in the current image, and to design visual controller are three key issues for achieving a robot visual servoing mission in natural scenes.

To solve the above problems, this paper proposes a robot visual servoing scheme based on the randomized tree classifier in natural scenes. The overall structure of the proposed scheme is shown in Fig. 1. The system mainly includes three parts as follows: building a randomized tree classifier, computing the current image feature, and designing the visual controller. The basic idea of the algorithm is: Firstly, user randomly selects an object to be manipulated prior to the visual servoing task execution. Then, the raw image about the object can be captured by the camera and used to generate a number of sample data sets for building the randomized tree classifier. Secondly, the current image features, which are represented as 2D pixel coordinates of the object image centroid, can be computed using the previously built classifier. Finally, visual control input can be calculated according to the image feature error and is applied to the robot to achieve the robot visual positioning for unknown objects. In Fig. 1, $\mathbf{f_d}$ represents the desired image features and $\mathbf{f}$ denotes the current image features (as shown by the red circle).

The basic principle of the three main parts of the servoing system will be stated individually in detail in next section.

4 Proposed Solution

In this section, detailed implementation of all functions highlighted in Fig. 1 will be described according to the order listed above, which includes the construction of the randomized tree classifier, the computation of the current image features, and the design of the visual controller.

4.1 Construction of the Randomized Tree Classifier

The main function of this module is to construct a randomized tree classifier for recognizing and detecting unknown objects randomly selected by the user. The whole process is shown in Fig. 2, which can be further divided into the following four steps.

Step 1. Selecting the unknown object to be manipulated. Firstly, the object selected randomly by the user will be put on the training station whose background is clean without any clutter, and the robot is also moved to the training station. Then, the desired image can be obtained through segmenting the current image captured by the camera mounted on the robot's end-effector. If the object is a 3D non-rigid body, its rough 3D model can be obtained through the ImageModeler software, and then the 3D object image will be obtained using the above method.

Step 2. Preprocessing. The main purpose of the preprocessing is to get the gray image of the desired image.

Step 3. Extracting the object features and generating the view sets. The stable affine invariant features are extracted on the preprocessed image. Then patches, whose center is each feature point, are obtained. These patches will form view sets of image feature points of the object, or object view sets for short.

Step 4. Establishing the randomized tree classifier. A randomized tree classifier can be built using these object view sets.

The detailed implementation of feature extraction, object view sets generation and randomized tree classifier establishment are described as follows.

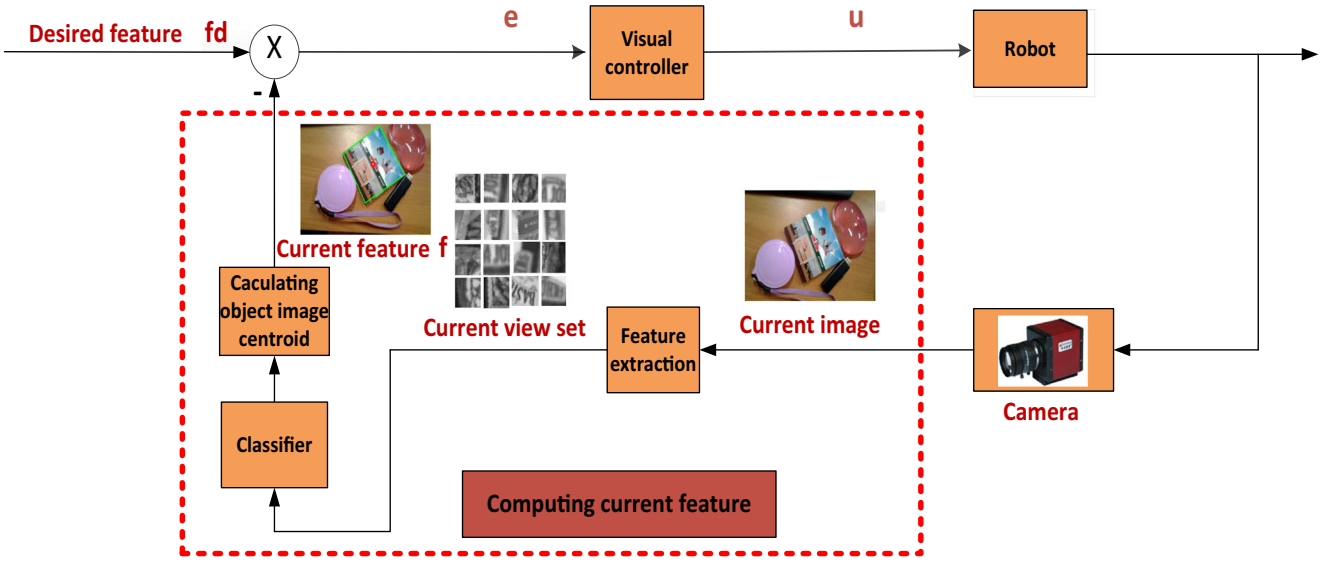


Fig. 1 Overall structure of the robot visual servoing system based on the randomized tree classifier

4.1.1 Feature Extraction

Characteristics of image features used in the control loop, especially for the visual servoing structure based on images, will directly affect the stability and robustness of a robot visual servoing system, and this is one of the important factors for robotic systems to be successfully applied to complex environments. Therefore, this paper adopts a two-level feature extraction method to obtain affine invariant stable features. Firstly, LOG-FAST feature extraction method is used for rapidly extracting features from the gray image of the raw object. The LOG operator is utilized to conduct Gaussian filtering and image sharpening processing so as to eliminate noises of the image. Then FAST-9 operator is used to extract the feature in different scale spaces.

To further obtain affine invariant stable features, it is necessary to conduct selection operation after the preliminary extraction. The main steps to achieve above task can be stated as follows:

Step 1. Generating M new images by conducting affine transformation on the grayscale image. Affine transformation is a combination between non-singular linear transformation matrix $\mathbf{A}=\mathbf{R}_\theta\mathbf{R}_\varphi^{-1}\mathbf{S}\mathbf{R}_\varphi$ and translation matrix $\mathbf{t} = [t_x, t_y]^\mathbf{T}$, where $\mathbf{R}_\theta$ and $\mathbf{R}_\varphi$ are the rotation matrices that correspond to θ and φ respectively, and they are within the ranges of $[-\pi, \pi]$. $\mathbf{S}$ = diag $\{\lambda_1, \lambda_2\}$ represents the image scaling transformation matrix, and λ_1, λ_2 are within the range of [0.2, 1.5]. Besides, t_x, t_y are within the range of [0, 2]. M new images are generated by randomly selected parameters θ, φ, λ_1, λ_2, t_x, t_y, and white noise is added into the generated M new images.

Step 2. Determining affine invariant stable features. Firstly, LOG-FAST feature extraction method is adopted to

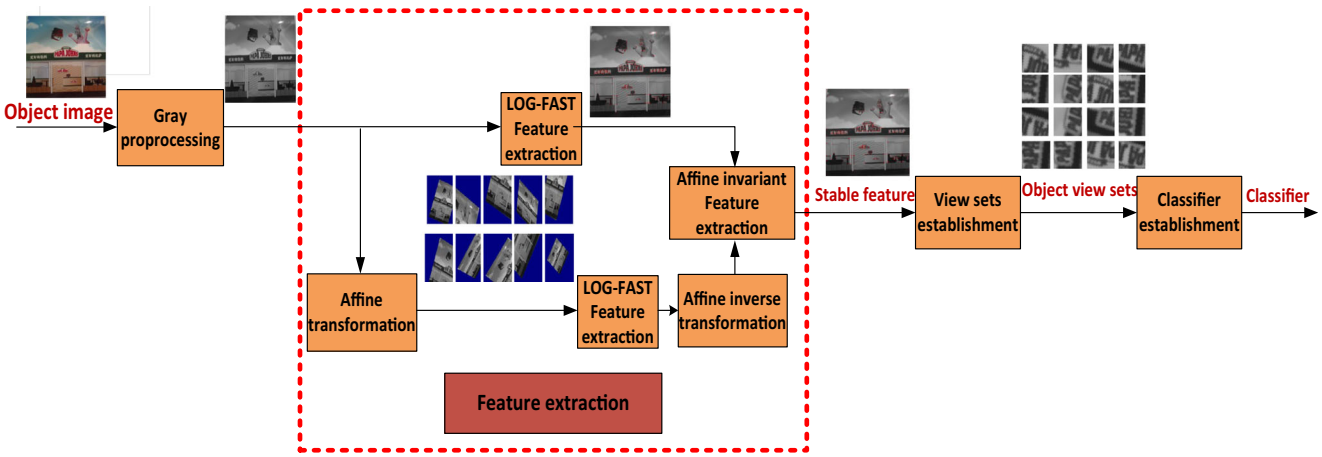


Fig. 2 Process of constructing a randomized tree classifier

extract the features on the M new images. Then, the inverse transformation is utilized to recover the extracted feature. Finally, the successful matching frequency between the recovered features and the features of the original image is calculated. The features with the top N frequency are considered as the "stable" features, which are illustrated with red circles in Fig. 3.

These features construct a feature set $\mathbf{K} = \{k_1, k_2, \ldots, k_N\}$, where 1 to N are ultimately determined stable features. Each feature is tagged to denote a class, and the different classes are indicated as $\mathbf{C} = \{c_1, c_2, \ldots, c_N\}$. Then these stable features are used to construct view sets to be employed for building the randomized tree classifier.

4.1.2 Establishment of View Sets

When capturing a frame of new image in the robot visual servoing process, the most critical problem is to determine whether the current image involves the object and object's location is in the image. The first step for solving the above problem is to detect the features on the object. So, the view set of each feature needs to be further established after the stable features are obtained. Feature patches are extracted on the M affine transformed images, and the size of each patch is 32*32. The view sets consist of all the extracted feature patches, and the size of view sets is M*N finally. All feature patches with the same number will form a small collection, thus for the N stable features, N small sets are constructed: $\mathbf{V}_n = \{v_{n1}, v_{n2}, \cdots, v_{nm}\}$, $1 \le n \le N$,where each $\mathbf{V}_n$ set includes m elements, and it is the view set of a feature. Figure 4 shows the view set of a certain feature, where different elements indicate the different locations for the same feature in different perspectives.

Fig. 3 Extracted stable features

4.1.3 Establishment of Randomized Tree Classifier

Binary decision tree is adopted in this system, which has only one root node and two child nodes. Each node is divided into two child nodes and the other nodes follow recursively until the bottom node has no branch. The bottom node is named as leaf node. The view sets are put into the root node, and patches of each view set traverse from the root to the leaf. In the traversal, every node will test the patch to determine whether this patch belongs to a certain node. This system adopts a gray image, thus the information gain of the gray information feature is the largest. Gray information is selected as the criterion of classification in this paper. The discriminant of each node can be written as:

$$T = \begin{cases} \text{left} & f(I(p,m_1) - I(p,m_2)) \ge \tau \\ \text{right} & otherwise \end{cases} \quad (1)$$

where τ is the default threshold, $I(p, m_1)$and $I(p, m_2)$ represent the gray values of two pixels which are randomly selected in the patch p entering the tree. When all randomly selected patches enter the randomized tree, the number of each class's patches entering the leaf node m and the number of all patches entering the leaf node M should be calculated. The ratio between m and M is the posteriori probability that a certain feature is identified as a certain class. If the number of the ith feature class's patches entering the leaf node is m_i and the number of all the patches entering the leaf node is M, then the posterior probability that the ith feature is identified as the ith feature class can be represented as:

$$P_{\eta(l,p)}(Y(p) = c) = \frac{m_i}{M} \quad (2)$$

where η represents the leaf node. The leaf node stores the posterior probability $P_{\eta(l,\ p)}(Y(p) = c)$ of each class.

In general, as the size of the sample set is large, it is difficult to guarantee a high accuracy result using one randomized tree classifier. Statistically speaking, more randomized tree classifiers can compensate this shortcoming. Therefore, the method of establishing more trees to speed up the process of the object detection is adopted. The next step is to use the built randomized tree classifier to detect objects and to compute the current image features on the current image captured by the camera in the natural scenes in real time.

4.2 Computing the Current Image Feature

The main function of this module is to detect the object in the current image and to compute the features of the current image. The whole process can be shown in Fig. 5.

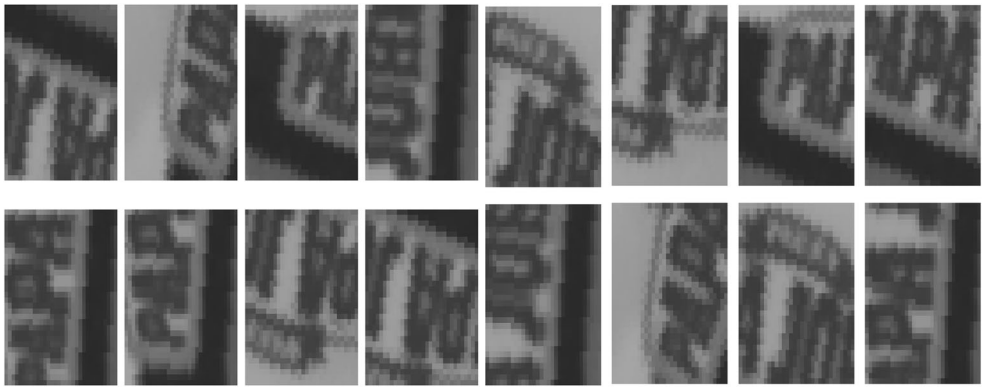


Fig. 4 View set of a certain feature

After a randomized tree classifier is established, one can obtain the posterior probability distributions of all classes of the features entering the leaves eventually $P(Y(X) = C|T = T_i, \text{ leaf } \eta)$ and the object view sets. When the current frame image comes during the process of the robot visual servoing, firstly one can extract features on the current frame image, and then select the neighborhood patches with a size of 32*32 for each feature. All obtained patches are used to establish the current view sets and enter the different randomized tree. Let $\widehat{Y}(p)$ be the estimated features label corresponding to the small patch p. Thus, $\widehat{Y}(p)$ can be considered by Eq. (3):

$$\hat{Y}(p) = \arg_c \max p_c(p) = \arg_c \left(\max \frac{1}{L} \sum_{l=1 \cdots L} P_{\eta(l,p)}(Y(p) = c) \right) > D_c \tag{3}$$

where $P_{\eta(l,\ p)}$ represents the posterior probability of the patch p entering the leaves nodes of the lth tree T_l, c is the feature class, $p_c(p)$ is the mean posterior probability of the class c and D_c is the default threshold (60% used in the later experiment). When the posterior probability is larger than D_c, the feature of this patch is classified as class c, and further, the center of this patch matches the feature of the class c in the desired image. Otherwise the patch belongs to the background or a misclassification and should be discarded. Thus, the match between the current features and the cth class features is implemented.

When the number of matching reaches a certain threshold, it is determined that the current image contains the object. In order to calculate the current image feature, the RANSAC algorithm is used to estimate the homograph matrix H. Then H is used to calculate the centroid coordinate of the object, which is the current image features. In Fig. 5, the centroid of an object image is represented by a red circle as shown in the picture named "current feature".

4.3 Design of the Visual Controller

After obtaining the current image features, one can further design the visual controller according to the image error to drive the robot for positioning an unknown object in natural scenes.

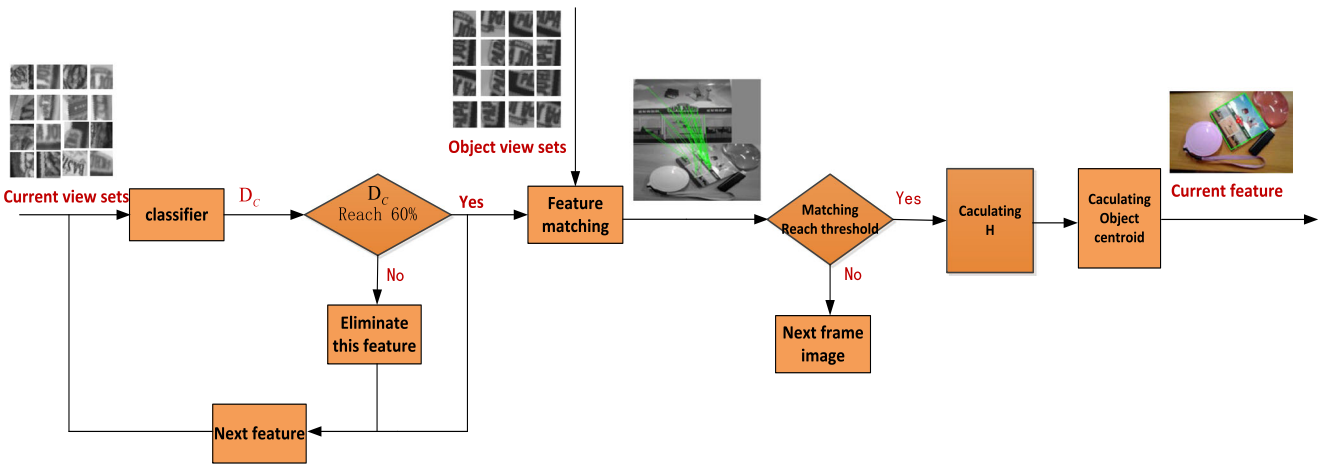


Fig. 5 Process of computing the current image features

The image error $\mathbf{e}$ can be defined as:

$$\mathbf{e}=\mathbf{f_d}-\mathbf{f} \tag{4}$$

where $\mathbf{f_d}$ and $\mathbf{f}$ are the desired image features and current image features, respectively.

For an image-based visual servoing structure used in the paper, image Jacobian matrix$\mathbf{J}_{\text{im}}$, as shown in Eq. (5), is often used to describe the relationship between the changes of the features in the image space$\dot{\mathbf{f}}$ and spatial velocity of the robot end-effector in the robot workspace $\dot{\mathbf{r}}$.

$$\dot{\mathbf{f}}=\mathbf{J}_{\text{im}}\dot{\mathbf{r}} \tag{5}$$

where $\dot{\mathbf{r}}$ is spatial velocity of the robot end-effector, including linear velocity and angular velocity. In other words, $\dot{\mathbf{r}}$ is the calculated visual control input u to be applied to the robot end-effector; $\dot{\mathbf{f}}$ can be viewed as the image feature error $\mathbf{e}$.

Then, a simple visual control law $\mathbf{u}$ based on the inverse Jacobian matrix can be designed as shown in Eq. (6).

$$\mathbf{u}=\begin{bmatrix}\upsilon\\ \omega\end{bmatrix}=\mathbf{K_p}\,\mathbf{J_{im}}^{+}\mathbf{e} \tag{6}$$

where $[\upsilon^{\mathbf{T}}\ \ \omega^{\mathbf{T}}]^{\mathbf{T}}$ is the camera's (or robot's) velocity, and $\mathbf{J_{im}}^{+}$is the pseudo inverse of image Jacobian matrix, and $\mathbf{K_p}$ is the proportional control gain. The Laypunov function is defined as $V = 1/2\mathbf{e}^{\mathbf{T}}\mathbf{e}$, $V > 0$. Then, the derivation of the Laypunov function $\dot{V}$ can be described as:

$$\dot{V} = \mathbf{e}^{\mathbf{T}}\dot{\mathbf{e}}=\mathbf{e}^{\mathbf{T}}\left(\dot{\mathbf{f}}_{\mathbf{d}}-\dot{\mathbf{f}}\right) \tag{7}$$

For eye-in-hand configuration used in the paper, camera is mounted on the robot end-effector and it moves along with the robot. So, in this camera-robot configuration, both robot visual positioning and visual tracking are considered as the positioning problem in the image features space. Therefore, $\dot{\mathbf{f}}_{\mathbf{d}}=\mathbf{0}$. Then,

$$\begin{aligned}\dot{V} &=\mathbf{e}^{\mathbf{T}}\dot{\mathbf{e}}=\mathbf{e}^{\mathbf{T}}\left(\dot{\mathbf{f}}_{\mathbf{d}}-\dot{\mathbf{f}}\right)=-\mathbf{e}^{\mathbf{T}}\dot{\mathbf{f}}=-\mathbf{e}^{\mathbf{T}}\mathbf{J}_{\mathbf{im}}\dot{\mathbf{r}}=-\mathbf{e}^{\mathbf{T}}\mathbf{J}_{\mathbf{im}}\mathbf{u}\\ &=-\mathbf{e}^{\mathbf{T}}\mathbf{J}_{\mathbf{im}}\mathbf{K}_{\mathbf{p}}\mathbf{J}_{\mathbf{im}}^{+}\mathbf{e}=-\mathbf{e}^{\mathbf{T}}\mathbf{K}_{\mathbf{p}}\mathbf{e}\end{aligned} \tag{8}$$

For robot visual control input in any ith directionu_i, it can easily prove that the Laypunov function satisfies $\dot{V}_i < 0$ by using Eq. (8). Therefore, the closed-loop system is asymptotically stable according to the Lyapunov stability theory, and the robot can be controlled to move to the desired image features.

5 Experimental Results

Five robot visual positioning experiments are conducted, mainly including visual servoing of 2D rigid objects and non-rigid objects in several natural scenes, to validate the performance of the proposed robot visual servoing scheme for unknown objects. Figure 6 shows the experimental platform for robot visual servoing, which contains a CCD camera, a MOTOMAN SV3 6-DOF manipulator robot and a personal computer with Windows operating system and the OpenCV 2.4.1 software. The CCD camera is mounted on the robot hand claw. It moves following with the motion of the robot hand claw, while its internal and external parameters are unknown. The image plane size is 1024*768 (unit: pixel), so the desired image feature $\mathbf{f_d}$ is the center of the image plane coordinate (512, 384) (unit: pixel). In order to evaluate the performance of the proposed robot visual servoing scheme in the real world, two types of five robotic visual positioning experiments were performed in the complex nature scenes, including unknown rigid objects and unknown non-rigid objects.

The visual control law shown in Eq. (6) is used, and the specific values of the relevant parameters are:$\boldsymbol{K_p}$ = *diag* (0.04,0.04).

The maximum joint running speed of Motoman SV3 6-DOF robot used in the experiment can reach 300°/s. In the experiment, for the sake of safety, the running speed of the robot is set at a low value of 10°/s. In Windows operating system, average image capturing and processing time per frame is about 0.156 s in the experiment, which can meet the real-time control requirements of the arm robot at low speed. The average processing time of the key components (taking the initial frame images captured in experiment 5.1.1 as an example, the average processing time of the key components for different initial frame images captured in each visual servoing task is basically unchanged) is as follows:

Fig. 6 Experimental platform

1). The processing time of the current image captured by the camera is about 0.0783 s. The average time per frame (per current image) captured by the camera is about 0.0160 s, and the average time of the feature extraction and feature matching are about 0.0531 s and 0.0092 s, respectively.
2). The sampling time of the controller is 1.35 s.

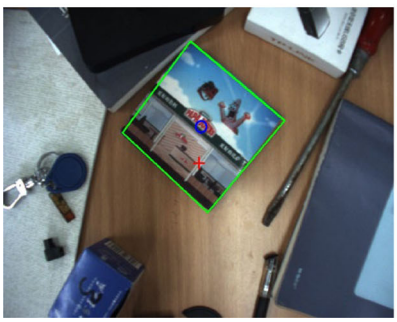
5.1 Robot Visual Servoing of 2D Rigid Objects in Natural Scenes

The purpose of this experiment is to verify the performance of the proposed approach for unknown 2D rigid objects in complex nature scenes, including cases with no-occlusion, with occlusion and with illumination change. The detailed results are provided in Experiments 1–3.

5.1.1 Visual Positioning of the Unknown Object without Occlusion

Visual positioning results are shown in Fig. 7. Fig. 7a illustrates the initial frame, where the center coordinate (512, 384) of image plane is visualized by a red cross. Namely, it is the desired image feature$\mathbf{f_d}$. The object is represented by a green rectangular bounding box. The object centroid is visualized by a blue circle and it represents the current image feature$\mathbf{f}(k) = (u, v)^T$. The positioning results of middle frame and end frame are illustrated in Fig. 7b and c, respectively.

(a) Initial frame


(b) Middle frame

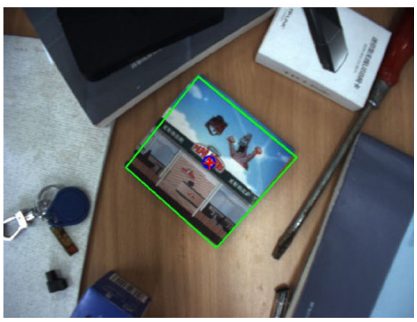

(c) End frame

Fig. 7 Visual positioning results without occlusion (2D rigid objects)

5.1.2 Visual Positioning of the Unknown Object with Occlusion

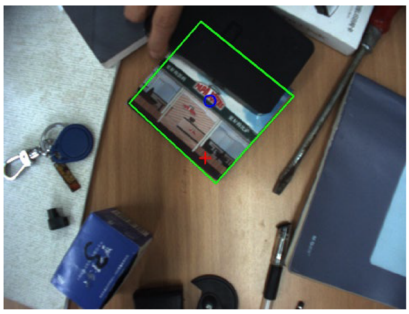
Visual positioning results are shown in Fig. 8. The positioning results of middle frame and end frame are illustrated in Fig. 8b and c, respectively. During the visual positioning, the object is occluded by the black Mobile Hard Disk. The experimental result shows that the robotic system still can detect the object and finish the positioning in spite of the large occlusion.

5.1.3 Visual Positioning with Illumination Variations

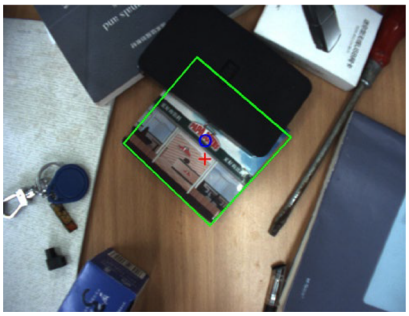
The purpose of this experiment is to verify the availability of the proposed approach under the scenarios containing overall and local illumination variations. Visual positioning results are shown in Fig. 9. Initial illumination is shown in Fig. 9a. During the visual positioning, the case of turning off the light in the robot workplace is shown in Fig. 9b, then add a beaming light as shown in Fig. 9c, then remove a beaming light as shown in Fig. 9d. Visual positioning results under the above three illumination variations are illustrated in Fig. 9 b and 9d, respectively. It can be seen from Fig. 9 that robot can successfully implement the visual servoing of unknown 2D objects in natural scene in spite of large illumination variations.

5.2 Robot Visual Servoing of Non-rigid Objects in Natural Scenes

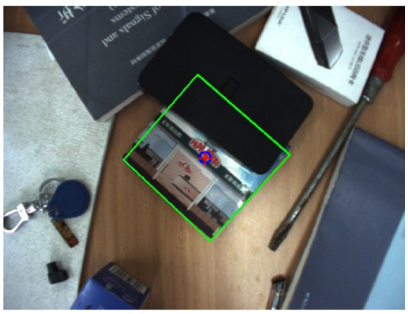
In order to further verify the effectiveness of the proposed approach for non-rigid objects in natural scenes, another two



(a) Initial frame



(b) Middle frame



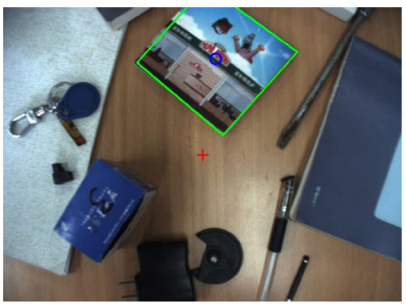
(c) End frame

Fig. 8 Visual positioning results with occlusion (2D rigid objects)

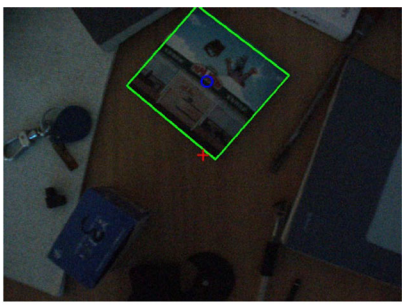
groups of visual positioning experiments were performed, including 2D non-rigid object and 3D non-rigid object. The detailed results are provided in Experiments 4–5.

5.2.1 Visual Positioning of 2D Non-rigid Objects

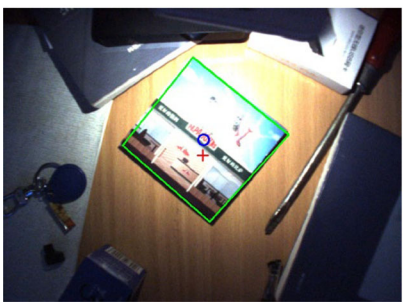
The purpose of this experiment is to verify the effectiveness of the proposed approach for 2D non-rigid objects in natural scenes. A sponge is selected as a 2D non-rigid object. The experiment is designed to simulate some practical applications for service robots, such as folding the clothes and operating the soft tissue in surgery and so on. During the visual positioning, the basic shape of the sponge has undergone tremendous changes. The sponge is performed with up and down mixture non-rigid deformation and visual positioning results are shown in Fig. 10. At the very beginning of the servoing, the sponge is squeezed up and its deformation is shown in Fig. 10a. Then sponge is squeezed down and visual



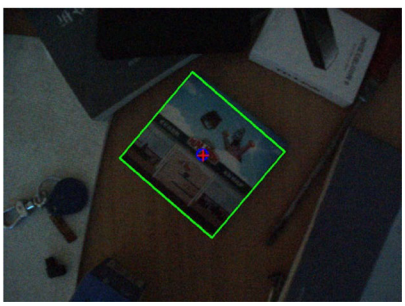
(a) Initial frame



(b) Middle frame 1: Turning off the light

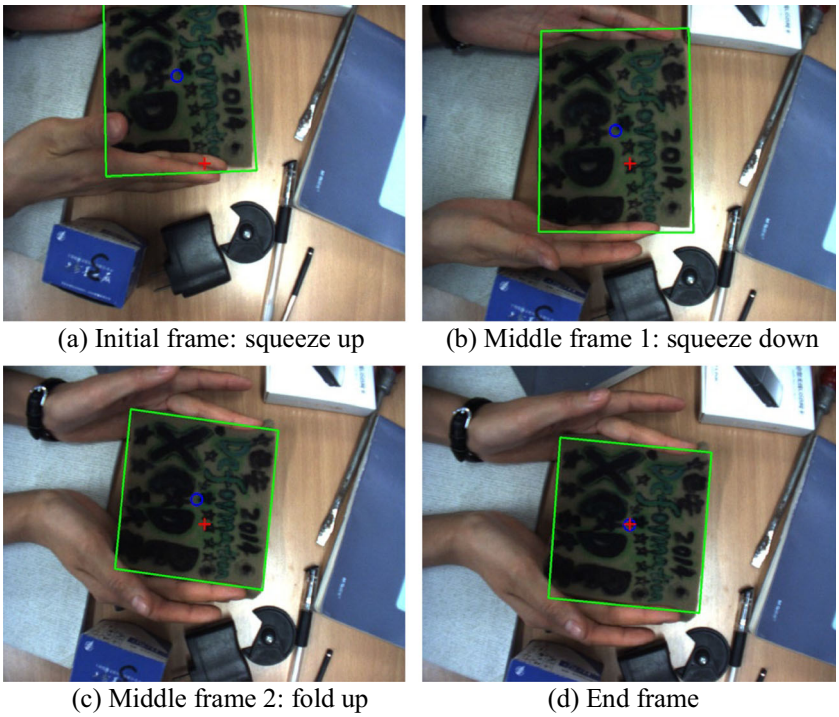


(c) Middle frame 2: Adding a beaming light



(d) End frame

Fig. 9 Visual positioning results with huge illumination variation (2D rigid objects)



(a) Initial frame: squeeze up (b) Middle frame 1: squeeze down

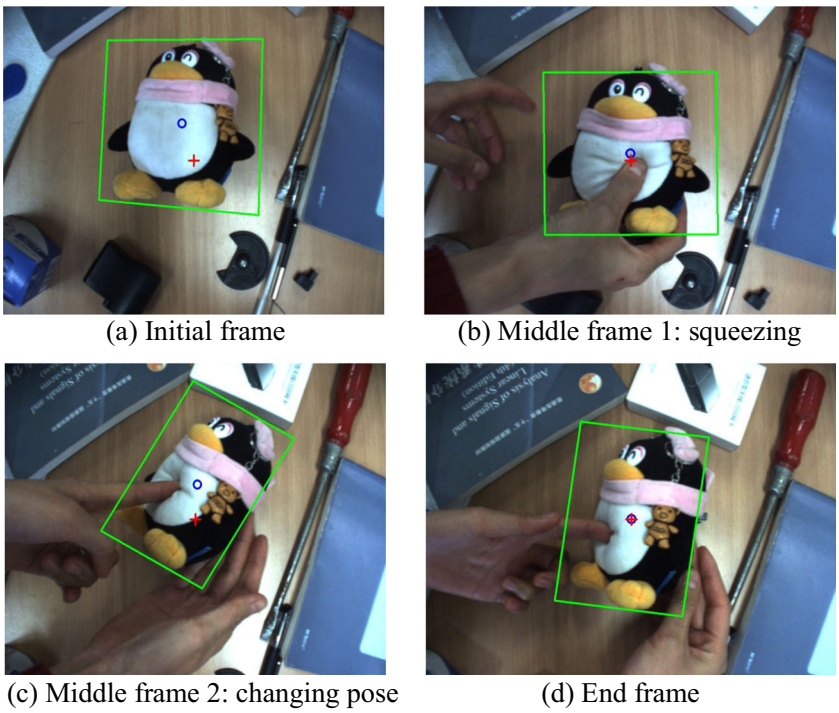
(c) Middle frame 2: fold up (d) End frame

Fig. 10 Robot visual positioning results with non-rigid deformation (2D non-rigid object-sponge)

positioning result is shown in Fig. 10b. After Fig. 10b, the sponge is folded up as shown in Fig. 10c until the end of visual positioning process as shown in Fig. 10d. It can be seen from Fig. 10 that the robot can still successfully implement the visual positioning of the 2D non-rigid object.

5.2.2 Visual Positioning of 3D Non-rigid Objects

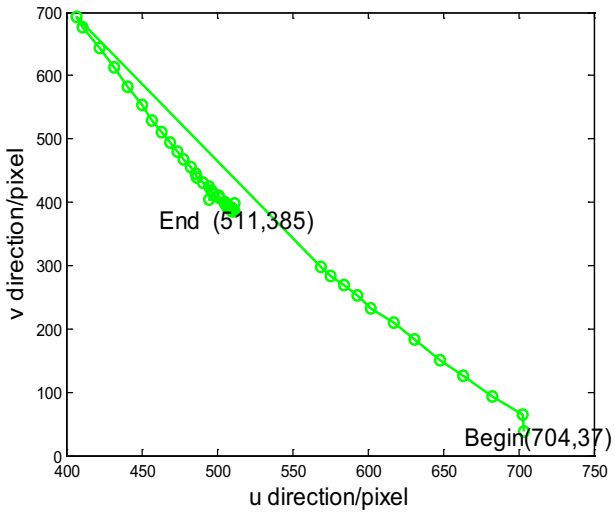
The purpose of this experiment is to verify the capability of the proposed approach for 3D non-rigid objects. A plush toy is selected as a 3D non-rigid object. As the plush toy has 3D



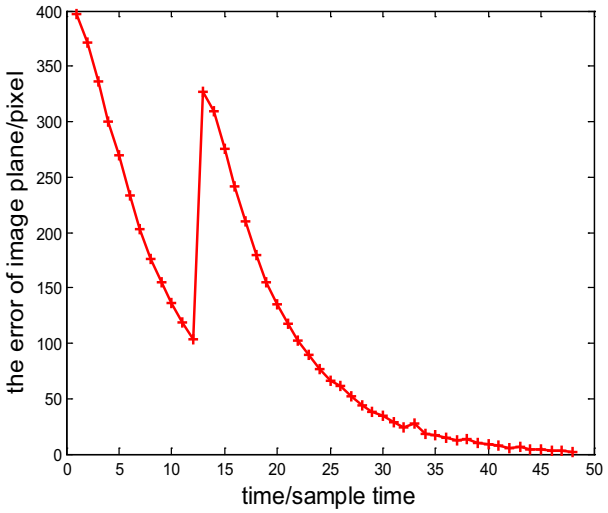
(a) Initial frame (b) Middle frame 1: squeezing

(c) Middle frame 2: changing pose (d) End frame

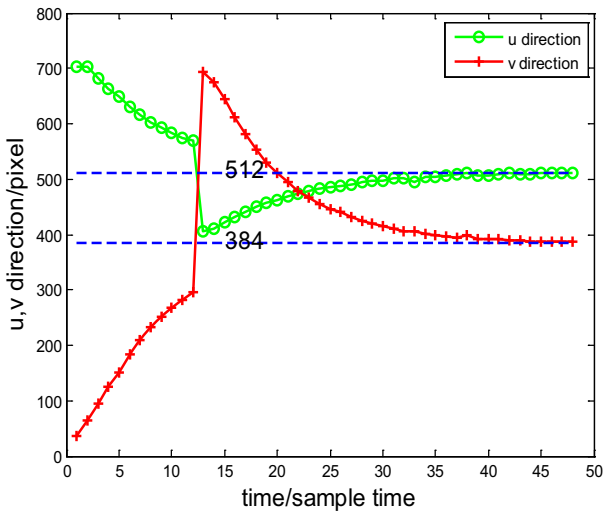
Fig. 11 Robot visual positioning results with large non-rigid deformation and pose variation (3D non-rigid object)



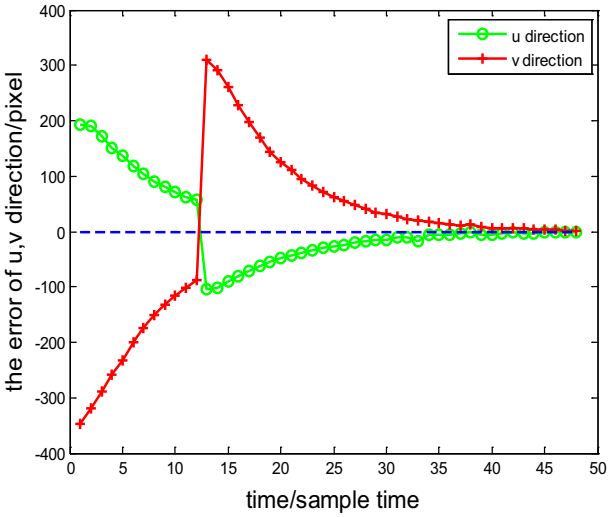
(a) Positioning trajectory in image plane



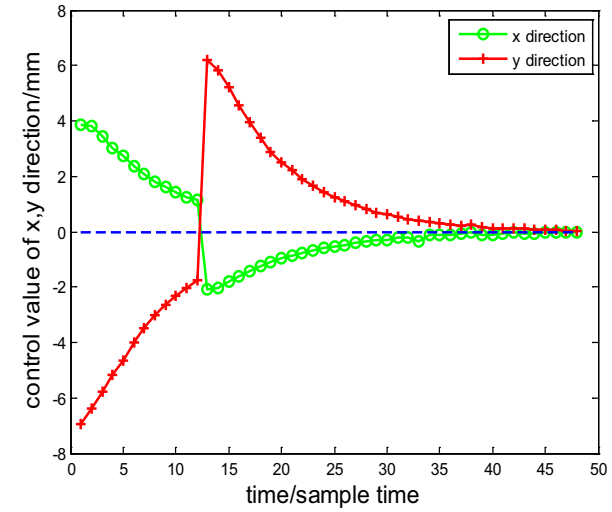
(b) Positioning error curve in image plane



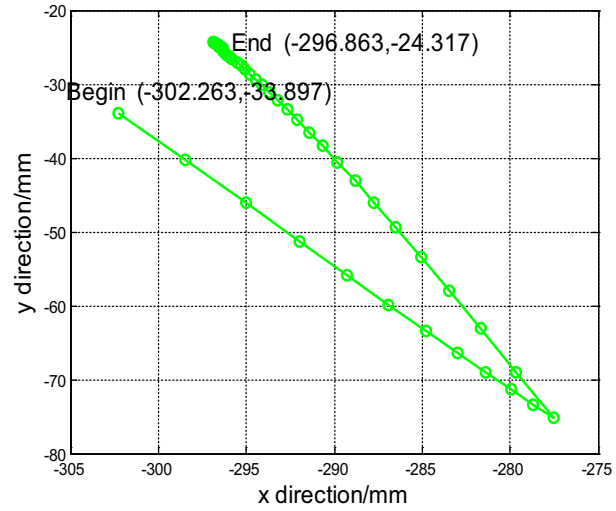
(c) Positioning trajectories in u and v directions



(d) Positioning error curves in u and v directions



(e) Control input curves in u and v directions



(f) Positioning trajectory in robot workspace

Fig. 12 Robot visual positioning results of the unknown object

characteristics, it is necessary to first build its rough 3D model by ImageModeler software using multiple views images of the plush toy during the visual positioning. The plush toy is then performed with certain non-rigid deformation and its pose is also changed. Visual positioning results are shown in Fig. 11. Initial pose and shape of the plush toy is shown in Fig. 11a and plush toy does not have any deformation. During the visual positioning, plush toy is squeezed and visual positioning result is shown in Fig. 11b. Then the pose of plush toy is changed and visual positioning result is shown in Fig. 11c. Continuing to be squeezed until the end of positioning as shown in Fig. 11d, it can be seen from Fig. 11 that the robot can still successfully implement the visual positioning of such a 3D non-rigid object.

5.3 Further Analysis and Discussion

Positioning trajectory curves in the different directions and different spaces can provide more insight into the visual servoing behavior. During the visual positioning, pose of the object can be changed freely. In this experiment, pose of the object is changed at the late servoing when the robot is nearby the desired position. Visual positioning results are shown in Fig. 12. The positioning trajectory of object is shown in Fig. 12a in the image plane u and v, and the positioning trajectory of robot is shown in Fig. 12f in the robot workspace x and y. Positioning error curves in the image plane, u and v directions are shown in Fig. 12b, c and d, respectively. It can be seen from Fig. 12 that the proposed visual servoing scheme has a better convergence performance.

Table 1 lists the mean, standard deviation (std) and max of the robot positioning error in 10 different poses of the 2D object without occlusion. The "max" represents the ratio between the max absolute of 10 times positioning error and the motion range of this direction. The model of CCD camera used in the experiment is MV-VS078FC-L, the image resolution is 1024*768, and the corresponding pixel size is 4.65μm*4.65 μm, that is, the physical size of each pixel in x and y directions is 0.00465 mm. It can be seen from Table 1 that the positioning error in x and y directions are about 0.004 mm and 0.048 mm respectively and the largest absolute relative error and standard deviation are also very small. Therefore, the developed positioning system has high accuracy with a positioning error within $e = \sqrt{e_x^2 + e_y^2} = \sqrt{0.004^2 + 0.048^2} = 0.05\text{mm}$.

As showed in Table 1, the five positioning experimental testing results show that the proposed approach of robot visual servoing based on randomized tree classifier can implement visual positioning of unknown objects in complex natural scenes with occlusion and illumination variations.

Table 1 Positioning error results of the object in 10 different poses (without occlusion)

error	mean/(mm)	std/(mm)	max/(mm)
x direction	−0.0040	0.0880	1.3283%
y direction	−0.0480	0.2189	1.8868%

6 Conclusion and Future Works

This paper proposes a robot visual servoing scheme to locate the robot to unknown objects in natural scenes with family services as application scenario. Five visual positioning experiments for unknown rigid object and non-rigid object in various nature scenes are conducted on a MOTOMAN-SV3X six degree-of-freedom manipulator robot. Experimental results show that the proposed scheme can effectively position an unknown object in complex natural scenes with strong robustness to occlusion and illumination variations and small positioning error within 0.05 mm. Furthermore, the system does not need any template nor any database of the object prior to the visual servoing task execution. Once the object is selected by the user freely, all the needed data can be obtained online and the robot can complete the positioning task automatically.

However, current method cannot position the unknown multiple objects at the same time. User can only specify one object for visual servoing task before performing the visual servoing task. If the current visual servoing task contains multiple targets, we have to reuse our method, which will become more time-consuming. In other words, our method is only suitable for the case where the number of targets is small and is not suitable for the case where the number of targets is too large.

As future works, the proposed robot visual servoing scheme will be further extended into two aspects: 1) to position the unknown multiple objects by combining multiple objects recognition and detection approaches [27, 28]; 2) to autonomously grasp the unknown object by combining deep reinforcement learning in view of its strong learning capability.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China under Grants 61833013, 61873200 and U20A20225, and the Natural Sciences and Engineering Research Council of Canada. The authors would like to express their sincere gratitude to the Editor-in-Chief, the Associate Editor, and all the anonymous reviewers whose insightful comments that have helped to improve the quality and presentation of this paper considerably.

References

1. Huebner, K., Welke, K., Przybylski, M., Vahrenkamp, N., Asfour, T., Kragic, D., Dillmann, R.: Grasping known objects with humanoid robots: a box-based approach. Proc. of the IEEE Int. Conf. on Adv Robot. (2009)
2. Rasolzadeh, B., Björkman, M., Huebner, K., Kragic, D.: An active vision system for detecting, fixating and manipulating objects in real world. Int. J. Robot. Res. **29**(2–3), 133–154 (2010)
3. Azad, P., Asfour, T., Dillmann, R.: Stereo-based 6D object localization for grasping with humanoid robot systems. Proc. of the IEEE/RSJ Int. Conf. on Intell Robots Sys (IROS). (2007)
4. Xin, J., Liu, D., Xu, Q.K.: LS-SVR-based robot uncalibrated 4DOF visual positioning. Contr Theory Appl. **27**(1), 77–85 (2010)
5. Hauck, A., Ruttinger, J., Sorg, M., Farber, G.: Visual determination of 3D grasping points on unknown objects with a binocular camera system. Proc. of the IEEE/RSJ Int. Conf. on Intell Robots Sys (IROS). 272–278 (2009)
6. Kim, J., Cauli, N., Vicente, P., Damas, B., Bernardino, A., Santo Victor, J., Cavallo, F.: Cleaning tasks knowledge transfer between heterogeneous robots: a deep learning approach. J. Intell. Robot. Syst. **98**(1), 191–205 (2019)
7. Reitelshöfer, S., Meister, S., Franke, J.: Recognition and description of unknown everyday objects by using an image based meta-search engine for service robots. Adv Engin Forum. **19**(2), 132–138 (2016)
8. Schiffer, S.: Integrating Qualitative Reasoning and Human-Robot Interaction for Domestic Service Robots. PhD Thesis, RWTH Aachen University (2015)
9. Zhao, Z.: Towards 3D Reconstruction and Semantic Mapping for Indoor Scenes. PhD Thesis. University of Science and Technology of China (2016)
10. Dune, C., Remazeilles, A., Marchand, E., Leroux, C.: Vision-based grasping of unknown objects to improve disabled people autonomy. Proc. of Robotics: Sci Syst. (2008)
11. Gratal, X., Romero, J., Bohg, J., Kragic, D.: Visual servoing on unknown objects. Mechatronics. **22**(4), 423–435 (2012)
12. Wang, H., Yang, B., Wang, J., Liang, X.W., Chen, W.D., Liu, Y.-H.: Adaptive visual servoing of contour features. IEEE/ASME Transac Mechatro. **23**(2), 811–822 (2018)
13. Wang, Z., Hirai, S.: Modeling and estimation of rheological properties of food products for manufacturing simulations. J. Food Eng. **102**(2), 136–144 (2011)
14. Higashimori, M., Yoshimoto, K., Kaneko, M.: Active shaping of an unknown rheological object based on deformation decomposition into elasticity and plasticity. Proc. IEEE Int. Conf. Robotics Automat. 5120–5126 (2010)
15. Shibata, M., Hirai, S.: Soft object manipulation by simultaneous control of motion and deformation. Proc. IEEE Int. Conf. Robotics Automat. 2460–2465 (2006)
16. Tokumoto, S., Hirai, S.: Deformation control of rheological food dough using a forming process model, in proc. IEEE Int. Conf. Robotics Automat. **2**, 1457–1464 (2002)
17. Das, J., Sarkar, N.: Autonomous shape control of a deformable object by multiple manipulators. J. Intell. Robot. Syst. **62**, 3–27 (2001)
18. Wu, H., Andersen, T.T., Andersen, N.A., Ravn, O.: Application of visual servoing for grasping and placing operation in slaughterhouse. Proc. of the Int. Conf. on Control, Autom Robot. 457–462 (2017)
19. Redmon, J., Angelova, A.: Real-time grasp detection using convolutional neural networks. Proc. of the IEEE Int. Conf. on Robot Autom (ICRA). 316–1322 (2015)
20. Zhong, X.G., Xu, M., Zhong, X.Y., Peng, X.F.: A robot grasping discrimination approach based on multimode feature deep learning. J Autom. **42**(7), 1022–1029 (2016)
21. Navarro-Alarcon, D., Liu, Y., Romero, J., Li, P.: Model Free Visually Served Deformation Control of Elastic Objects by Robot Manipulators. IEEE Trans. Robot. (2013)
22. Navarro-Alarcon, D., Liu, Y., Romero, J., Li, P.: Visually served deformation control by robot manipulators. Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA). **21**(1), 5239–5244 (2013)
23. Berenson,D.: Manipulation of deformable objects without modeling and simulating deformation, IEEE/RSJ Int. Conf. on Intell Robots Syst (2013)
24. Berry, F., Martinet, P., Gallice, I.: Real time visual servoing around a complex object. IEICE Trans. Inf. Syst. **83**(7), 1358–1368 (2000)
25. Ma, Y., Liu, X., Zhang, J., et.al. Robotic grasping and alignment for small size components assembly based on visual servoing. Int. J. Adv. Manuf. Technol. 106(11–12):4827–4843 (2020)
26. Li, J.A., Xie, H., Ma, R., Low, K.H.: Output feedback image-based visual servoing of rotorcrafts. J. Intell. Robot. Syst. **93**(1), 277–287 (2019)
27. Ren, S., Girshick, R., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transac Pattern Anal & Mach Intell. **39**(6), 1137–1149 (2017)
28. Yin, Y., Li, H., Fu, W.: Faster-YOLO: an accurate and faster object detection method. Digit Signal Proc. **102**(7), 102756 (2020)

Jing Xin received the B.S., M.S., and Ph.D. degrees from the Xi'an University of Technology, Xi'an, China, in 1997, 2003 and 2007, respectively. She is currently a Professor with the Key Laboratory of Shaanxi Province for Complex System Control and Intelligent Information Processing, Xi'an University of Technology, Xi'an, China. In 2010–2011, she held a one year visiting scholar position at the Australian Centre for Field Robotics, University of Sydney. In 2012 and 2016, she held a three-month visiting professor position at the University of Technology, Sydney, respectively. She is currently a Senior Member of the Chinese Association of Automation (CAA), a Member of the Committee of the Chinese Association of Automation Construction Robot. Her current research interests include manipulator robot visual servoing, mobile robot visual navigation, and robust object tracking.

Caixia Dong is currently a M.S. candidate in control theory and control engineering with the Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing at Xi'an University of Technology, Xi'an, China. Her research interests focus mainly on manipulator robot visual servoing.

Youmin Zhang received the B.S., M.S., and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 1983, 1986, and 1995, respectively. He is currently a Professor with the Department of Mechanical, Industrial and Aerospace Engineering and the Concordia Institute of Aerospace Design and Innovation, Concordia University, Montreal, Quebec, Canada. His main research interests include fault detection and diagnosis (FDD), fault-tolerant control (FTC), fault-tolerant cooperative control (FTCC) of single and multiple unmanned aerial/space/ground/marine vehicles, smart grids, and applications of unmanned systems to forest fires, power lines, environment, natural resources and disasters monitoring, detection, and protection by combining with remote sensing techniques. He has authored 8 books, over 550 journal and conference papers, and book chapters. Dr. Zhang is a Fellow of CSME, a Senior Member of AIAA and IEEE, President of International Society of Intelligent Unmanned Systems, and a member of the Technical Committee for several scientific societies. He has been an Editorial Board Member, Editor-in-Chief, Editor-at-Large, Editor or Associate Editor of several international journals, including as a Board Member of Governors and Regional Representative of North America of the Journal of Intelligent & Robotic Systems. He has served as the General Chair, Program Chair, and IPC Member of several international conferences, including as a General Chair of the 5th Int. Symp. on Autonomous Systems (ISAS), Hangzhou, Dec. 17–19, 2021 (www.isas-conference.com), and 2022 Int. Conf. on Unmanned Aircraft Systems (ICUAS) to be held at Dubrovnik, Croatia during June 21–24, 2022 (http://www.uasconferences.com/). More detailed information can be found at http://users.encs.concordia.ca/~ymzhang/.

Yumeng Yao is currently a M.S. candidate in control theory and control engineering with the Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing at Xi'an University of Technology, Xi'an, China. Her research interests focus mainly on manipulator robot visual servoing.

Ailing Gong is currently a M.S. candidate in control theory and control engineering with the Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing at Xi'an University of Technology, Xi'an, China. Her research interests focus mainly on manipulator robot visual servoing.